FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Simulating Driver-Pedestrian Interaction and Intentions Inference

Rui Pedro Correia Soares

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: João Tiago Pinheiro Neto Jacob, PhD

Co-Supervisor: Rosaldo José Fernandes Rossetti, PhD

July 12, 2019

Simulating Driver-Pedestrian Interaction and Intentions Inference

Rui Pedro Correia Soares

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: José Manuel de Magalhães Cruz, PhD
External Examiner: Vítor Hugo Mendes da Costa Carvalho, PhD
Supervisor: João Tiago Pinheiro Neto Jacob, PhD

July 12, 2019

Abstract

The appearance of autonomous vehicles (AVs) has become a crucial and ever more emerging topic, giving way to new discoveries in the behavioral aspect of robotics. It is imperative that AVs know how to interact with pedestrians during driving contexts. Interaction between drivers (either active or passive ones) and pedestrians is something crucial to driving experiences in intersections, signaled or non-signaled crossings. If these situations are not well controlled and interaction outcomes are not predicted, risky situations occur.

Thus, the creation of models pertaining to such crossing scenarios is a powerful tool to infer over real-life ones, where testing hypotheses may be too dangerous.

When creating these models one needs to take into consideration the importance of extrinsic and intrinsic factors of driver-pedestrian interactions. These latter ones, ie. the communication between the two, can give drivers useful hints of intentions by pedestrians. Knowing such intentions means knowing how to react to pedestrian crossing scenarios. Studies found that non-verbal communication is very useful in asserting confidence by pedestrians.

Related work is extensive. Many authors have already used factors like vehicle speed, among many other factors to build predictive models that can output confidence of the pedestrian crossing. Some authors used neural network approaches to estimate pedestrian poses in crossing scenarios.

If AVs need to predict such intentions in driving contexts, predictive models that take into consideration this communication must be built. Thus, the contributions that this dissertation brings are twofold.

Firstly, the study of the relevance of non-verbal communication when predicting outcomes of crossing scenarios. It is important to justify the need or lack thereof of computing non-verbal communication and using it along other factors in predictive models.

Secondly, the creation of a methodology to better study which factors influence driver-pedestrian interactions, and how information about them can be extracted directly from the source into results. These results can be used for predictive models to be employed in simulations.

The solution was elaborated using a dataset created from a set of experiments and some image segmentation techniques.

These experiments led subjects to drive around in a virtual cockpit in an environment filled with pedestrians. Their actions during pedestrian encounters permitted the analysis and extraction of conclusions.

Obtained results emphasize the need of knowing how to react to non-verbal communication alongside the normally used metrics to predict pedestrian crossings. They relate the main factors of these driver-pedestrian interactions and provide better knowledge about their influence in the outcomes of such scenarios.

Resumo

O aparecimento da temática do veículos autónomos (VAs) tem tornado cada vez mais este tópico um tema crucial para discussão, facilitando o aparecimento de novas descobertas no lado comportamental da robótica. É imperativo que os VAs saibam interagir com peões em contextos de condução. Interação esta entre condutores (ativos ou passivos) e peões é crucial em experiências de condução, sejam elas em cruzamentos, passadeiras ou mesmo fora de passadeiras. Se estes cenários não forem bem controlados e os resultados das interações não previstos, há possibilidade de situações de perigo para os envolvidos.

Assim, a criação de modelos relativamente a estes cenários de atravessar a estrada revela-se uma ferramenta poderosa para inferir sobre cenários reais, one testar hipóteses pode ser custoso ou perigoso. Ao criar-se estes modelos, é preciso ter em consideração a importância dos fatores extrínsecos bem como intrínsecos das interações entre condutores e peões. Estes últimos, que se traduzem como a comunicação estabelecida entre os intervenientes, podem dar pistas importantes para saber prever os resultados de cenários de atravessar a estrada. Estudos concluíram que comunicação não-verbal entre peões em condutores é muito útil para que os peões possam demonstrar a sua confiança em atravessar.

Trabalho relacionado com este tema é extenso. Vários autores já concluíram que fatores como a velocidade de veículos, a densidade do trânsito, bem como muitos outros fatores são relevantes na construção de modelos preditivos neste contexto. Alguns autores usaram metodologias baseadas em redes neuronais para estimar poses dos peões ao chegarem à berma da estrada. Como VAs precisam de prever tais intenções de atravessar, modelos preditivos que tomam em consideração esta comunicação não-verbal terão de ser construídos, Portanto, as contribuições que esta dissertação trará são múltiplas.

Primeiramente, o estudo da relevância da comunicação não verbal ao prever resultados de cenários de travessia. É importante saber justificar a necessidade ou a falta desta aquando da computação de comunicação não verbal, bem como a sua utilização paralelamente a outros fatores das interações. Segundamente, a criação de uma metodologia para melhor compreender quais os factores que influenciam interações entre condutores e peões, e como a informação destes pode ser extraída diretamente do acontecimento e traduzida em dados. Estes dados são usados para modelos preditivos que serão empregues em simulações.

A solução foi elaborada usando um banco de dados angariado a partir de experiências e algumas técnicas de segmentação de imagens, entre outras. Estas experiências consistiram nos participantes a conduzir num *cockpit* virtual, num ambiente repleto de peões. As suas ações durante os confrontos com peões permitiram a análise e extração de conclusões.

Os resultados obtidos enfatizam a necessidade de saber reagir à comunicação não-verbal de peões conjuntamente com outros fatores já estudados, na predição de travessias. Estes relacionam os fatores pricipais dentro das interações entre condutores e peões, e permitem um conhecimento mais profundo sobre a sua influência nos resultados destes cenários.

Acknowledgements

This thesis could not have been completed without the help of some people.

I would like to thank my supervisor João Jacob for helping from the start in delving in this thesis' theme. And also, for the feedback throughout. I would thank professor Rosaldo Rossetti for providing humor and great help during the first part of this project.

I would especially thank my friends for all the support, all the coffees and Red Bull, all the music and all the fun that allowed us to relax and unwind. They are (in no particular order) Rita Torres, Miguel Silva, Diogo Cepa, João Soares, Rafaela Fernandes, Jonas Loureiro, Bruno Marques, Gonçalo Ribeiro, Nuno Cr, Rita Lima and many others. Thank you for all the hours spent at the lab or at the piano, doing some covers. They meant a lot.

Special thanks to everyone that chose to participate in the experiments and those that helped me when the first batch of experiments did not go well. I couldn't have kept it together without your help.

I'd also like to thank my family for always providing me with support and loving words. Thank you mom, my godmother and thank you to my twin, João. I couldn't have done this without you all.

"Let me not then die ingloriously and without a struggle, but let me first do some great thing that shall be told among men hereafter."

Homer

Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial Neural Network
AV	Autonomous Vehicle
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
LIDAR	Light Detection and Ranging
MAS	Multi-Agent System
ML	Machine Learning
NN	Neural Network
RFID	Radio Frequency Identifier
SVM	Support Vector Machine
VR	Virtual Reality

Chapter 1

Introduction

With the appearance of the potential deployment of Autonomous Vehicles (AVs) in normal traffic settings alongside other vehicles comes a need to understand how they should react to typical traffic events. Be those interacting with other vehicles, traffic lights or pedestrians willing to cross the road, every possible behavior that can be taken place by the AV should be well understood and explained.

1.1 Context

When introduced in traffic, a vehicle and its driver navigate the environment to reach a destination while also ensuring no accidents throughout the trip. When the driver is passive inside an AV, full control of the vehicle's response is transferred to the AV itself. What this means is that the vehicle has full responsibility in following traffic rules, reacting accordingly to pedestrian crossings and detouring in case of potential accident.

Such trust that is given to the AV should be well-founded. Thus, managing interactions is a crucial part of the vehicle's viability.

To investigate more about this context, authors can resort to simulations to study it in a controlled and manageable environment. However, simulating intention inference and driver-pedestrian interactions is not a trivial task. Authors have been somewhat successful in predicting actions of pedestrians in real-life settings. However, when translating such interactions to models no single methodology is determined to be the optimal one. Many different approaches have been explored, but comparison of them is due. It is important to define a main model and factors to use in predicting pedestrians' intentions.

1.2 Motivation

Given such a domain, events may be fast-paced and potentially dangerous for interacting drivers and pedestrians. In normal crossing events, these interactions may simply be quick eye contact or halting, for instance. Nevertheless, a whole plethora of potential behaviors to be taken place by all participants in interaction should be accounted for.

This justifies such a need for understanding the vehicles' reasoning. One should study these by modelling interactions where the environment is easily translatable to real-life scenarios while also diminishing the potential danger for anyone involved. Various ways of modelling have been identified, although there isn't a defined consensus on which is the most useful in this context.

The surfacing of Serious Games as a mechanism to study serious situations in a game-focused environment, allowing users to play a game while also contributing to its translation into real-life scenarios can be useful in this context. Given a driving simulator, players could help understand their interaction with virtual pedestrians [RAKG13]. This ensures no risk is present during the study, while such environments allow complex interactions with the primary driving task to be assessed [GRJ+14, AGR+13, GRO12].

From there the AV should be able to predict interactions' outcomes so as to ensure the best possible reaction in each one.

It is imperative that interactions are well studied and that no risk is present for populations if AVs are to predict human intentions. Being an unsolved problem, its resolution would guarantee the minimization of misunderstandings in potential crossing scenarios, as well as the possibility of better commuting experiences for everyone. Furthermore, it would present a basis to further research of AVs when in contact with pedestrians. Predictive techniques are to be applied there on after, and a controlled simulation serves as a testbed for model verification and extracting conclusions.

1.3 Problem

We should strive for seamless commuting scenarios and general understanding of both pedestrian and drivers' intentions.

If drivers cannot infer what pedestrians will to do during the driving experience, risky scenarios may rise. Furthermore, they need to do this in a way that does not cause misunderstandings or impasses.

The current state of the art in this context allows for the prediction of crossing outcomes in general real-life observed scenarios. But they do not bring knowledge on how to transfer these predictions into simulated environments for both improvement and testing in safe environments.

One should strive for the use of a inference methodology that allows for results within this scope in a simulated environments.

This problem is thus divided into three parts.

- **RQ1**: In regards to analysis of real-life scenarios, how can one extract knowledge from driver-pedestrian interactions? And what factors need to be taken into consideration within them? And how is this done?
- **RQ2**: Studying these factors should be done in controlled scenarios, so as not to generate risky scenarios. How can such modelling be done?
- **RQ3**: How should drivers react to human non-verbal communication in crossing scenarios? And how can they infer over their intentions?

1.4 Goals

The goals of this dissertation pertain to the problem at hand. These are:

- Study and compare methodologies to extract knowledge from Driver-Pedestrian interactions (**RQ1**)
- Analyzing such interactions in a model, to visualize them in a controlled and non-risky environment (**RQ2**)
- Obtain a basis for a predictive model over pedestrian intentions (**RQ3**).

1.5 Expected Contributions

The dissertation's methodology will provide at the end a pipeline architecture to serve as the basis for a tool to predict human intentions during crossing scenarios. This pipeline will be able to provide some insight into pedestrians' willingness to cross the road, as well as the main factors that influence this willingness.

1.6 Thesis Outline

This thesis is structured as follows: Chapter 2 provides a state-of-art review of literature on driver-pedestrian interaction. Chapter 3 defines the solution's methodological approach to the problem and its implementation. Chapter 4 goes through how experiments based on the methodology were approached, and discusses results gathered through them. Chapter 5 explains the conclusions that were obtained and proposes some future work based on such conclusions.

Chapter 2

State of the Art

Some factors to be taken into consideration when analyzing driver-pedestrian interaction are the differences in the internal and external characteristics of each participant in the interaction. Vehicles may be of very different sizes or shapes, and depending on the place where the crossing takes place vehicles may be followed by a line of traffic or grouped with others in multi-lane crossings. On the other side, pedestrians are quite heterogeneous as well: they may be of any age, size or gender. They may also cross in groups or individually. Moreover, their decision making and intentions are not predictable, so each participant needs to be analyzed if one wants to fully understand a scenario.

It is important to note that when a driver approaches a pedestrian, their communication is also not predictable. Some drivers may choose to stick to eye contact (the most common method) [RKT17], perform hand gestures, flash their lights, or use verbal communication to signal the pedestrian to cross [Šu14]. Others let their vehicle's speed be the main factor that lets a pedestrian measure if it is safe to cross. Naturally, when seeing an approaching vehicle, a pedestrian expects it to break before the zebra crossing. The drivers are evidently the ones that make the judgment of the vehicle's yield when approach a zebra crossing. Some drivers may choose to yield long before the crossing. Others, on the other hand, may even speed up to assert their reluctance to yielding [Var98]. Pedestrians need to make sure they can estimate the approaching vehicle's speed correctly [SZW+15].

Pedestrians are also capable of asserting their intentions. Either by performing eye contact with the driver, signaling them or taking an aggressive stance and crossing despite unsafe situations, they just as well have responsibility on the outcome of the interaction [SR11]. It is relevant to note that cultural differences might play a role in these happenings [CPD+18].

It is important to take in consideration all these factors that play a role in crossing scenarios. One simple distraction or miscommunication between the participants means a potentially dangerous scenario for anyone involved [Ris85], even more so if the vehicle is autonomous and the

driver is passive. Pedestrians aren't yet familiarized with autonomous vehicles, and this leaves situations open for misinterpretation, especially since humans don't have constant mental states and patience for interpretation [Wol16], although studies have attempted to bridge this communication gap [HL+18].

2.1 Data Collection Techniques

When studying a domain like this document's, it is desired that real-life events are well translated into a new constrained domain where one can analyze them without taking into consideration every potential external factor that falls outside of the scope of the study. Building this new domain means having to select all relevant variables that alter the domain in a way. From there, one can build a viable model where the study is done in a controlled environment.

Pedestrian behaviour has successfully relied on data collection techniques in the past [ARF+14, ARJ+17], allowing for the study of the decision-making process of these actors.

In this case, modelling driver-pedestrian interactions requires understanding of the vehicle's driving, the pedestrian's crossing, the physical environment where the interaction takes place and all communication that takes place between the entities [RKT17].

Depending on the scope of the study, previous work has mostly funnelled on either one of the factors. That is, studies generally focus on how the vehicle's driving may affect pedestrian crossing, or vice versa. Despite the study's approach to the problem or the sub-problem that is explored, previous work was usually developed by gathering data for modelling using one of four approaches: by using observational techniques, instrumented vehicles in driving scenarios, by simulation environments [MSSE15] or even by questionnaire-based work. Either one has its advantages and assumptions which lead to inevitable knowledge gaps. Nevertheless, they are all able to translate interactions to be used in some model. This section aims at understanding each one, along with some other approaches.

2.1.1 Observational Techniques

Perhaps the most immediate way to gather data about interactions would be to observe it directly in the real world. This makes sense, since it is in essence a behavioral study. Crossing scenarios are everywhere: they are a crucial part of vehicle scenarios that involve some sort of driving. Be them at signaled or non-signaled zones, pedestrians need to cross a road at some point when commuting by foot and must pay attention to vehicles coming their way. They are an essential part of commuting, and one that can be influenced by many factors [RKT17] [KV13].

The participants' heterogeneity makes scenarios a complex sequence of actions. Each participant's intentions and decision making need to be observed and justified if one attempts at using an observational study. However, this task is not trivial. Drivers and pedestrians employ different strategies based on their responsibility on the outcome of the scenario, and are acted on based mainly on the participant's confidence, assertiveness or feeling of safety [RKT17]. Thus, observing a scenario is observing a choice of strategies derived from the interaction's internal

and external factors. Externally, traffic density, the presence of traffic lights [BZSMM13], fast approaching vehicles and their collision times [LD15] [SAW17], groups of pedestrians waiting to cross, road properties and weather [Šu14] [CGM+18], along many others are the main attributes to be noted to study the interaction. Internally, the communication taking place between participants, be it implicit or explicit (or none at all) should be accounted for.

To extract these conditions while observing interactions the tool of choice is the use of video recordings [MSSE15]. Typically, data is recorded in long sessions in a particular crossing. This allows for the creation of rather large datasets to be employed in the study. Some studies focus on a particular crossing and focus their study on that specific case. Others record in several crossing locations. The recorded data is then processed either by automated methods or by using human experts as the judges of what took place in each instance of interaction.

Besides video recordings, some studies enrich their study data by using sensors of various kinds (RFID, LIDAR [SR11], etc.) directly on the field. Such sensors are mainly used to gather information about the participants' speeds and trajectories.

Focusing on the aforementioned factors that influence each scenario, human experts are the ones to decide how each one played out. Such a task is rather time-consuming and tedious, as well as prone to human error. Nevertheless, humans are typically good at interpreting such scenarios, since they have experience in them. Depending on the scope of the study, some variables may be of greater importance to define than others, and thus should be more sensitive to misinterpretations.

Pedestrian tracking and interpretation of scenarios may also be done automatically by use of computer vision and clustering techniques. Such clustering has been studied to be useful when analyzing pedestrian flow and their microscopic behaviors [Hoo02].

When doing a comparison study of different models (gap acceptance vs. motorist yield), some authors [SUBTW02] decided to use an observational study to gather data on accepted and rejected gaps as well as motorists' yield in a particular crossing scenario. They justified their choice based on the location's traffic flow, the different gaps that could be observed therein, the absence of obstructions as well as it being an uncontrolled mid-block crossing (which the study focused on). These results were similar to those of a study where the authors also recorded a particular crossing to study requirements for simulation [GCVB18] (Fig. 2.1).

Other authors analyzed how road safety measures would affect irregular pedestrian crossings [MS16]. Using observational techniques followed by descriptive studies, they concluded that traffic lights and refuge islands affect how pedestrians act the most.

When developing a dataset using observed data in crossing scenarios, authors concluded with confidence that the majority of pedestrians glance at the driver when crossing. This underlines the importance of communication between them, since it is so prevalent [RKT17].

2.1.2 Instrumented Vehicles

If the study's focus is on the driver's side, some metrics regarding the driving need to be analyzed. Of course, like mentioned before, vehicular speed can be estimated using sensors placed on the

Figure 2.1: The observed location in an observational study [GCVB18]. Research was only conducted on this particular location.

field. Nevertheless, these do not provide information about the vehicle's workings, such as acceleration, breaking or turning as a mechanical act. Moreover, further information about the driver could be required. It might be important to analyze their bio-metrics, like the heart rate, exudation or brain activity, for example.

To provide such data, a vehicle can be instrumented to gather driver and vehicle actions during the experience. They allow to easily plot out each action against the driving time, and visualize actions that took place at any instant [LT06]. A strategy in a crossing behavior can be visualized by this, and compared to other runs.

This kind of collection technique makes no room for human error or subjectivity. Thus, the absence of bias makes for the presence of clean data. This hastens the analysis process and makes it easier to directly compare two different interactions and what took place. It does not, however, make for the direct observation of scenarios. Implicit communication between the driver and the pedestrian will not be gathered by this technique, which raises a requirement to couple it with another one. Nevertheless, it provides a way to obtain a large dataset of a driver's actions.

In order to develop a dataset of driving footage coupled with a notation to indicate a goal in turns or causes for breaking, some researches used an instrumented vehicle to gather sensor data to use in the development of the notation. Although not focused on driver-pedestrian behavior, the study shows how an instrumented vehicle can provide valuable data [RCMS18], as well as displaying the usefulness of computer vision techniques for detection and tracking.

2.1.3 Simulation

Recently though, the use of agent-based modelling in traffic simulation have evolved to the concept of Artificial Transportation Systems (ATS), with different applications [RLT11, RL15]. Some techniques followed the approach of observing driver-pedestrian behavior in real-life settings, resorting to the so-called naturalistic data. Although useful to observe evidently credible scenarios, those studies are not controlled in regards to risk. Crossing scenarios may be dangerous for any of the participants, more so if the vehicle is autonomous and the pedestrian does not know how to react. Thus, this emphasizes a need for controlling such environments, such as ATS and simulation.

One can do this by using computer simulated environments where the potential risk of interactions is naught. The environments are closed and have no external factors that may alter the

Figure 2.2: The realistic weather simulation achieved in a simulation study [DRC+17].

results of scenarios. Thus, they are useful to investigate interactions since they provide a good comparison to real scenarios. However, this is not without disadvantages. To build a simulation, one needs to build it based on a set of assumptions that will be the basis of simulation. Simulations may be static and simply built on a set of rules that will be taken into account when extracting conclusions. They can also be dynamic and change throughout development, either by programming or user input.

Driving simulators, along with other vehicle simulators have long been used to study transport scenarios virtually. The user is immersed in a virtual cockpit where he must navigate the vehicle in a setting, with a trajectory. The user faces no risk of danger even if the simulated environment is dangerous [YS14]. In fact, it allows for the preparation of the user in regards to such scenarios.

In addition to this, driving simulators may be a fun experience for the user, while having him learn the simulated skill. The emergence of Serious Games has made it possible to create environments that provide a fun experience for users while simultaneously providing knowledge that can be extracted from their plays. Serious Games have been used to study simulation in the context of transport settings, and although validation for this approach regarding Driver-Pedestrian interaction has not yet been achieved, it can be a valuable tool to be used in knowledge extraction [RAKG13].

Many authors have studied potential accidents in driving using simulation data [SNL03] [NS04]. In their study of black spots in traffic, some authors [WSB15] decided to run a simulated environment of a complex crossing scenario. No input is provided to the system (it is a closed MAS). Nevertheless, they were able to easily visualize interactions and state that the main factors that influence an interaction for the driver were vehicle size, visibility of the crossing, and time to yield. Other authors have studied how pedestrians affected traffic using simulation [BK09].

Authors created a study simulation tool [DRC+17] (Fig. 2.2) that allows for the visualization of AVs in a town scenario. It does not focus on driver-pedestrian interaction directly, although it can be visualized as a simple obstacle avoidance process. The simulation achieved is flexible and allows to analyze some external factors in driving, eg. the weather.

2.1.4 Other Types of Techniques

Although the three previous approaches tend to be the main ones that authors stick to when researching driver-pedestrian interaction, some other ways of collecting relevant data exist. Among them, the main one that can be used is the approach of using interviews and questionnaires to obtain direct input from participants.

Posing questions about people's driving or usual commuting habits provides an opinion-based dataset that may provide interesting information [Dí02][HFJS06]. Depending on the scope of the study, these questions may be focused on things like:

- Demographics, like age, gender, handicaps, etc.
- Macroscopic factors in interactions, like whether they travel in groups, if they usually drive/-commute alone, if they avoid traffic or big intersections, etc.
- Microscopic factors in interactions, such as whether they tend to look at the other interaction's participants, if they gesture at them, if they say thanks, if they usually yield, etc.
- More subjective data, such as whether they tend to blame the other participant in an accident, if they think a particular demographic tends to have more accidents, etc.
- Given a particular scenario as example, if they think the example's participants' actions were justified, or an error, lapse, violation, etc.
- Other data like the amount of accidents they have been a part of, or road rules that they typically avoid, etc.

A questionnaire provides a similar set of behavioral answers as to an observational study. It comes as a cheaper and safer alternative (since no behaviors are actually performed whatsoever) [DSD+17]. This type of information is useful because it allows for comparison between the answers and the actual habits in the same scenarios. People tend to soften their responses if it incriminates their behavior, so answers are commonly biased [DSD+17]. Thus, a questionnaire is not a good tool to be used by itself to analyze the scope, due to its subjective nature. It is better to couple it with other data, typically observational, and extract conclusions from both supports.

To the best of my knowledge, not many studies rely solely on interviews to draw conclusions. Since most studies in this area are behavioral, it is natural to wish to observe the scenarios, even if the interview's answers cover most of the needed variables for the study. This what was done in a study regarding AVs in Mexico City [CPD+18], where the authors coupled observed data with a post-study questionnaire. It provided some info on their perspective of the study and helped draw conclusions regarding the AV.

2.1.5 Comparison

Choosing a data observation technique depends on the scope of the problem. Some problems may require direct visualization of behaviors, while others need as a primary requirement the negation of risk when collecting data.

Of course, one is not limited to employing just one technique [MSSE15]. Like mentioned before, several techniques can be coupled to make up for the flaws of others. Observational studies allow for direct visualization of interactions, but make room for human error. Instrumented vehicles allow for precise data collection of driver metrics, but are expensive and do not provide much by the way of visualizing interactions. Continuing this comparison, each method's advantages and disadvantages can be summarized as below:

Table 2.1: Data Collection Techniques' Advantages

Data Collection Technique	Advantages
Observational Study	Not much setup required
	Allows for direct behavior observation
	Allows to visualize communication between participants
	If equipped with sensors, allows to view macroscopic properties of interactions
	Generates big datasets
	Generally cheap and easy to setup in multiple scenarios
Instrumented Vehicles	Requires proper equipment
	Collects microscopic driver data
Simulation	Depending on the system, allows for complete observation
	Allows for study of every scenario, including potentially risky ones
Questionnaires & Interviews	Direct input from interaction's participants
	Potentially large sample size

Table 2.2: Data Collection Techniques' Disadvantages

Data Collection Technique	Disadvantages
Observational Study	Prone to subjectivity and bias
	Macroscopic properties require additional equipment
	Analysis requires a lot of time and consistency
	Needs proper setup for correct visualization
Instrumented Vehicles	Generally not cheap
	Needs proper equipment
	Interactions are not directly analyzed
Simulation	Setup is extensive
	Conclusions may not directly translate into real life ones
	Behavioral Observation depends on assumptions and setup
Questionnaires & Interviews	Highly subjective
	Not very valuable if nothing to compare to
	Conclusions need to take into account sample used

Generally speaking, and taking into consideration the information in tables 2.1 and 2.2, one should employ a combination of observational studies with questionnaires to enrich the data gathered through observation. This combination is very effective for comparison of behaviors and for direct visualization of instances of communication between drivers and pedestrians. However, it does not easily provide good data about the driver's point of view.

If one is willing to analyze further into the driver's POV, then an instrumented vehicle coupled with observational studies or questionnaires should be used. Driver data will be collected through instrumentation, while behavioral data can be gathered through observation. It does require coordination of the two, and it is not without cost.

When the focus of the study is repeated instances of behavior and controllability while reducing risk, simulation is the main tool to be employed. It is worth noting that a combination of all these should be the method to gather the most possible data about the domain: an instrumented simulation where a user drives through several scenarios that can be observed from a pedestrian's point of view would be maximizing potential data while minimizing risk.

2.2 Modelling of Driver-Pedestrian Interaction

Perhaps the most interesting part of the domain at hand is how it can be studied through the construction of models. These models attempt at representing the domain: variables like microscopic and macroscopic factors of interaction, as well as the process and outcomes of interactions themselves should be translated into the model.

From the moment where proximity between drivers and pedestrians is achieved and participants are aware of one another, intentions are communicated [LD15]. Here, assertiveness and confidence are tested [CGM+18]. It is up to the pedestrian to accept a gap to cross. These are factors that occur during the arrival of the driver. When the pedestrian is able to cross, whether dangerously or not, communication is still made be it by gestures or the act of crossing itself [Šu14]. All these factors may be taken into consideration when building a model to simulate interaction.

Gathered datasets usually have quite a large size. To make sense of the information therein, data needs to be prepared and organized in order to be applied in some modelling technique. Depending on the used gathering technique, data may have random noise or irrelevant attributes that must be suppressed before feeding it to the model.

The application of a model lets generalize conclusions extracted from the sample to the whole population. Since conclusions are directly derived from the model, its choice and fitting must be justified if conclusions are to be accepted. The approach requires also that the model is adequately evaluated, for the same reason [Kit17]. Many performance measures exist, and it is up to the author to pick which ones make the most sense to use on the each application.

When validating models, authors mainly test gathered metrics against known statistics. Two main approaches of modelling interactions arise. One is a statistical or support modelling that simply aims at outputting a set of relationships between attributes in the data. The other is a completely different approach where the whole scenario is developed in a simulation using a Multi-Agent System (MAS). If agents are modelled correctly, one can extract metrics from the simulation itself. It evidently implies the use of a simulation technique to gather data.

2.2.1 Modelling using Support and Statistical Tools

The most common approach to modeling this domain is the use of statistical models (sometimes coupled with ML approaches) to analyze data. These types of models are useful in the sense that they provide descriptive analysis of the domain at hand. From gathered data, they compile information and relate attributes in data to infer relationships between them. Observation of attributes in this way is done mostly numerically, with a discussion of results done after the methodology.

Evidently, statistical analysis of results is crucial to their understanding. Methods for statistical analysis are well-known and theoretically validated. Thus, this allows such conclusions to be validated if the confidence in them is high enough and they are not manipulated through the presence of a weak sample, many outliers or noisy data (although the latter two can be compensated for).

In the scope of driver-pedestrian interaction, many studies take gathered data and try to find correlations between factors that interfere with interactions. Mainly used are descriptive and predictive analysis of results (even though a predictive analysis is usually preceded by a descriptive one and does not provide visualization of behaviors).

Descriptive approaches categorize populations and find common characteristics between subgroups within them. Generally speaking, descriptive approaches are what lets us understand the data and its statistics. They let observations be summarized in graphs or tables. Thus, they simplify approaching data for either results discussions or to be followed by predictive analysis [DHGRH80]. In this context, they allow us to visualize the factors to take into consideration when observing interaction.

Many studies describe how external factors such as traffic density [GCVB18], pedestrian flow [SSL05] or vehicle speed influence different reactions in pedestrians. Some authors developed a descriptive analysis of encounters at zebra crossing scenarios [Var98], to reveal relations between the car's arrival speed, the car's arriving time and the pedestrian's arriving time. Some profiles of driver behavior surged when using this approach. Other authors added to those relations that even though many pedestrians see a car incoming, they choose to cross anyway [SMM], increasing the risk of a dangerous interaction. When comparing the same population regarding traffic light presence and the presence of traffic law violations, authors noticed that traffic lights do not seem to reduce tendency to commit violations, except when pedestrians are in big groups or traffic lights are countdown lights [BZSMM13][WW10]. Using aggregated observation data and correlating driver yield behavior with pedestrians' attempts to cross, some authors [KV13] stated that such behavior influences pedestrians' own behavior in crossing. It is worthy of note that cultural changes have an influence in how pedestrians react to drivers [CPD+18] (compared samples from different geographic locations). All these authors used a descriptive methodology to explore the empirical data (mainly observation) in research.

2.2.2 Agent Modelling

Perhaps the most intuitive way of visualizing behaviors in a model should be to have entities reenact behaviors themselves. This can be done if one develops an agent approach. By creating sets of rules that agents (pedestrian and driver agents, mainly) follow, one can achieve a viable method to visualize behaviors. Many MAS systems have been developed to visualize what drivers take into consideration in traffic and crossing scenarios. From these systems statistics can be extracted, to have an even better understanding of the behaviors at hand.

MAS systems are usually implemented in three-dimensional simulations. Whether static (that is, user input plays no role in changing the environment) or dynamic environments (for example, driving simulators), agents usually have some sort of sensing of their environment and of themselves.

It is important to note that although MAS systems allow for great visualization of behaviors, they ultimately have one big setback. That is, MAS systems are only as good as the rules which the agents follow. Simplified rules offer simplified behavior. Although this may be desirable, it

does not provide input into edge cases or specific interaction contexts, which may belong to the scope of study.

Studies using MAS systems outline the importance of traffic density and vision obstruction in driver's behavior [WSB15]. Drivers that cannot see pedestrians tend to act more erratically when entering interactions.

Other authors outline the importance of pedestrian gap acceptance in agent simulations [SAW17]. They concluded that using normal distributions to replicate vehicle speeds in agent systems was the one that provided the least errors. Knowing vehicle speed before anticipating a gap where pedestrians could cross was crucial to replicate good pedestrian behavior.

2.2.3 Other Types of Modelling

Despite the previous two being the most common methods of observing behavior in driver-pedestrian interactions, other methods have also been employed by authors in attempt to study this domain.

One method is the creation of a game theory model to replicate interactions as a game theory problem [FC+18] [CCB+18]. After defining a set of states and the outcome of such states, interactions can be visualized in every possible state, providing completeness in visualization.

Other methods include, for example, cellular automata models.

2.2.4 Comparison

Choosing a modelling technique depends on the type of visualization one is aiming for. Both modelling techniques presented provide insight into driver-pedestrian interaction.

Table 2.3: Modelling Techniques' Advantages

Modelling Technique	Advantages
Statistical & Descriptive Modelling	Relatively simple to develop
	Many types of models to choose from
	Provides validated insight into correlated attributes
Agent Modelling	Provides direct visualization of behaviors
	Intuitive interpretation of scenarios
	Depending on the system, may be highly adjustable
	May accept user input and perspective

Table 2.4: Modelling Techniques' Disadvantages

Modelling Technique	Disadvantages
Statistical & Descriptive Modelling	Does not provide direct visualization of behaviors
	Interpretation might be vague or non-intuitive
Agent Modelling	Requires set of rules
	Scenarios' specificity depends on rule system
	Setup is not simple
	Depending on system, may be too generalized

Information from tables 2.3 and 2.4 provides insight into the two main techniques' advantages and disadvantages. While Statistical and Descriptive Modelling works well in relating factors that go into the interactions, visualization of such interactions is simply done via graphing and plotting. Thus, they provide indirect visualization of interactions but direct observation of correlations and causality of those factors. On the other hand, agent modelling provides direct visualization of behaviors in rather human-friendly way. However, they require extensive setup and creation of a rule set that allows for such interactions. Typically statistical modelling is performed over the observed results in MAS, and compared to those observed in real life scenarios.

2.3 Pedestrian Behavior Inference

From exploring data and describing it in descriptive approaches, authors should want to obtain predictions from results obtained in such exploration. If attributes present relationships of correlation between each other, they might be useful to build a model that predicts the outcome of the combination of those attributes. Although fallible, some authors reach a reasonably good degree of success when predicting results. This degree of success should take into consideration the usual performance measure metrics for models, based on the confusion matrix concept. Of course, reaching a good success rate in test data is not very useful if such success is due to factors like bias and variance, that affect fitting of such model.

Some models that can be applied to predictive problems include, to name a few, Decision Trees (and other tree-based methods), Linear Regression (and Multiple Linear Regression), Support Vector Machines, Logistic Regressions and Neural-Network based approaches. Of these, only a few have been used for behavior prediction in pedestrian contexts.

Decision Trees are a type of supervised predictive model that can be used for classification or regression tasks [RM14]. After processing a training set to build the model, it describes a strategy to be employed as a tree, with different levels (nodes) and edges between them (splits). They allow to predict the outcome of a strategy represented by the path chosen for data in a testing set and output this outcome as a quantitative (regression) or qualitative (classification) value. This is the most common prediction method employed by authors to visualize associations and rules that dictate behaviors in interactions, because of its simple visualization and construction. It is not,

however robust to outliers or noisy data, so data should be prepared accordingly if one wishes to predict behaviors using this methodology.

Using decision trees, some authors found that visual awareness of each participant is one of the most important factors in a predictive approach [CGM+18]. When pedestrians are aware of the incoming vehicle, they tend to avoid risky situations. The most influential explanatory variables are indeed driver and pedestrian behavior prior to contact, like other authors confirm [Chu12]. Other transport scenarios have been studied with decision trees as well [ZLT16].

Support Vector Machines divide a training set into two different labels by fitting a linear function that describes the variable that influences them. It can be used for non-linear variables and for multi-labeling scenarios, although the approach needs to be adjusted. They are models that are quite robust to outliers in data. SVMs have not been thoroughly employed in driver-pedestrian interaction analysis or transport domains, although some authors have attempted to employ their use in studies [LLZX08].

Neural Network approaches make use of Artificial Neural Networks (ANNs) to predict a set of numerical outputs from a set of inputs. Sets of connected nodes organized in layers communicate with the next layers through weighted edges, and calculate an output based on such weights and the input given. Output values are adjusted using backpropagation algorithms [EREHJW86] [Wer06]. Although results using this approach have been obtained, mainly in the computer vision domain [HS06], its theoretical background is yet to be further defined. Neural Network approaches may include the use of standard ANNs, or Deep NNs , like Convolutional Neural Networks (CNNs).

Work on pedestrian detection and prediction in crossing scenarios has been explored, mainly using deep networks.

Convolutional Neural Networks are mainly used in for segmentation and feature extraction in image recognition tasks [SZ14] [GDDM13]. They are different from NNs because of their use of convolutional and pooling layers, as well as their use of ReLU to correct negative values in convolutional layers. After a series of combinations of convolutional layers and pooling layers (combinations depend on the architecture used), classification tasks are done through the use of a normal fully-connected NN.

Authors were able to track pedestrians and detect their regions in images using CNN approaches [LTWT14] [YLW16] [FL18]. Architectures differ in every approach.

Pedestrian tracking in images is an emerging topic that has been studied successfully, though not always in crossing scenarios. The current state of the art is able to identify pedestrians in crowded scenarios with precision, using tracking by detection. Bounding boxes for each proposed pedestrian region are detected and then tracked for future frames. Although these methods take a lot of data and time to process results, these results are usually high-precision.

One approach is the development of Switchable Deep Networks [LTWT14] (coupled with Switchable Restricted Boltzmann Machines as switchable convolutional layers) to simultaneously detect hierarchical features of pedestrians' body parts in images, while also attributing them semantic meaning. It showed promising results in such semantic classification and separation from

background clutter. However, despite being an excellent detection tool, it does not support pedestrian tracking.

A different approach is the use of skeleton model fitting in detected objects using CNNs [FL18] (Fig. 2.3). This allows to detect pedestrians' poses in each image of the dataset. The authors also made sure to guarantee pedestrian tracking using a multiple object tracking-by-detection methodology. They made a reference regarding the limitation of detection in non-tracking scenarios. Nevertheless, this CNN technique to detect poses was connected to a SVM classifier (instead of the usual fully connected layers) to identify crossing behaviors during driver-pedestrian interactions. Thus, their approach is very interesting since it makes full use of CNNs and predictive classifiers to identify pedestrian intentions using poses.

Figure 2.3: Pedestrian Tracking and pose estimation [FL18]

Besides using these predictive models, other authors employ other approaches to the driver-pedestrian interaction prediction problem [DT17]. Some authors built a framework for Deep Reinforcement Learning (along with many other algorithms for comparison) for AVs, based on three essential tasks like prediction, recognition and planning of behaviors [SAPY17]. Although results were not very specific, their work provides insight into autonomous driving tasks. Other efforts have nonetheless suggested the use of appropriate simulation environments to test with the different abilities of autonomous vehicles to interact with their surrounding environment [FRBR09, PR12].

2.4 Summary

Data collection is a crucial part of a study about driver-pedestrian interaction. Interactions are complex and have many factors that lead to a successful crossing or a dangerous setting. Thus, when taking on this domain, one needs to negotiate what data is crucial to the study. Each data collection technique has its own way of generating a dataset of behavior or microscopic or macroscopic communication instances.

Users taking on this domain should pick a combination of methodologies in order to maximize the amount of relevant data for the scope of their study. Nevertheless, observational techniques are the ones most frequently used due to their evidently realistic results.

After gathering data and compiling it into a dataset, visualization of such interactions therein is needed. Thus, datasets are fed into visualization models. Each one provides different insight to take into consideration when analyzing driver-pedestrian scenarios. While some methods focus on relating factors in interactions, others simply allow for controlled visualization.

Modelling behaviors allows authors to visualize them in controlled scenarios, and although some provide direct visualization, other provide interesting metrics to take into consideration in this scope.

Predicting behaviors in behavior models is a hard task and not a thoroughly explored area. Authors have been doing it mainly through observing behaviors in datasets. From there they have correlated some important attributes that factor into crossing outcomes. Work on this using pose estimation and tracking has also been done recently.

Chapter 3

Methodological Approach

In previous chapters a discussion of recent work in data collection, modelling and pedestrian behavior inference was provided. This discussion was necessary to introduce the methodological approach for inferring pedestrian behavior in crossing scenarios. The approach focused mainly on the solution for the data collection and modelling problems, leaving the behavior prediction for future work.

The solution was developed using two main factors in mind, namely, what types of visual information and contextual information about pedestrians influence the behavior of drivers. Drivers that are able to make the most out of available information are thus able to make better judgment of pedestrians' intentions.

3.1 Environment

Since the scope of the project focuses on simulating real-life scenarios, a trustworthy and stable environment for development was needed. Such an environment was one that allowed for visualization of driver-pedestrian interactions. These visualizations were necessarily able to be broken down into segments for analysis, thus bringing up any contextual information relevant to interactions.

An existing open-world simulation of an urban environment was used. This simulation and virtual environment fall under a project named SIMUSAFE, a project whose aim is to improve the current state of the art in driving and traffic simulation technology. This simulation developed in Unity3D, is composed of a large environment depicting a city, filled with realistic buildings, roads, sidewalks, crosswalks, and more. Moreover, it is highly modular, and individual parts of it are able to be disabled or switched for different ones. This flexibility leaves development in any module to be as less dependable on other modules as possible.

The city features many different urban scenarios, all in one model. Some areas are much more urban and tight for circulation (Fig. 3.1), while others are more open and empty (Fig. 3.2). They contain two, three or four-way intersections where crosswalks are at each end. Sidewalks vary in size and distance from the road. In many roads is busier areas, parked cars line the sides of the

Figure 3.1: Environment buildings, roads and other objects.

roads, providing an element of business to the scenarios, while still feeling realistic and not out of place. Roads also vary in size, some having more than two lanes and some being only one way. Some roads also end in roundabouts, where the environment feels much more open and dynamic. Such differences in areas were taken in consideration for the study of pedestrians in crosswalks. Interactions naturally differ in more crowded areas of the city.

Besides the static objects comprising the city scenario itself, pedestrians are able to be inserted in the simulation as well. These pedestrians are in a module of their own. Pedestrians in the environment behave as agents, and have a source of origin that can be set. A prefab of the model (Fig. 3.3) is then instanced there. Each pedestrian is able to collide with other elements of the simulation as well as any other pedestrian. Their movement is based on way points scattered around relevant corners and intersections. Their knowledge of the city layout is defined by the graph of such way points. Each pedestrian agent has a certain goal to get to at all times, and strives to the best of its ability to reach this point. This is done while also taking in consideration the costs of navigating in sidewalks or roads. Pedestrians should evidently feel more inclined to navigate in spaces that are allocated for their movement, ie. crosswalks and sidewalks. Thus, their movement tends to keep to those places. However, if navigating towards their goal is done much easier even if crossing in places not designated for this, they are able to do so as well. This flexibility in their movement means that their crossing behaviors are not one-dimensional, and they cannot be expected to follow the rules at all times. This is similar to how humans tend to cross the road and the thought process they go through in such navigation.

In their way to reach a goal, pedestrian agents receive perceptions that influence their behavior. It is necessary that the agent knows some stimuli of the world that surrounds it in order to achieve a higher degree of realism. The fact that they are surrounded by other non-static agents like themselves means that they should be able to react to other agents' intentions. Agents will strive to dodge others in their path and not to interfere with others' paths. Since they sometimes share

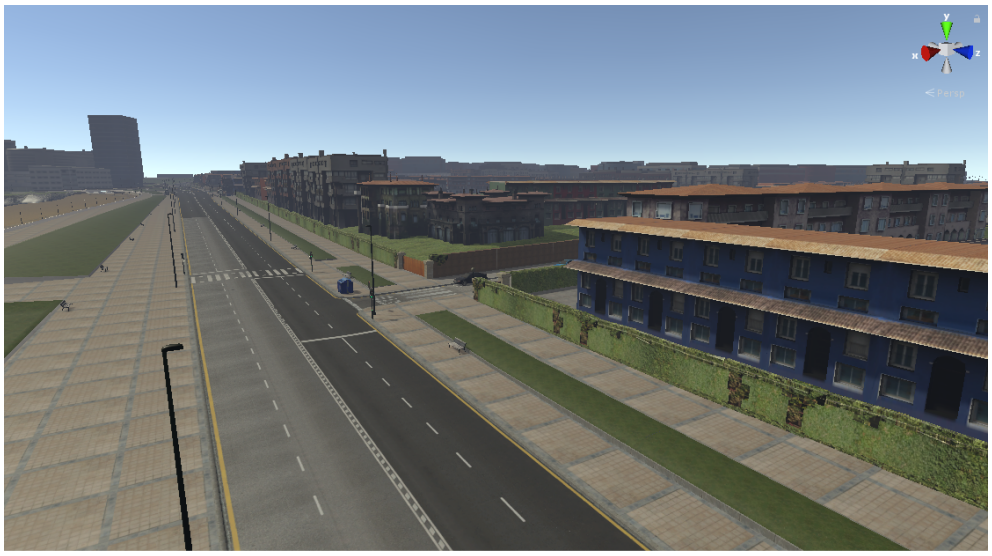


Figure 3.2: Another part of the environment.

the same space, on a sidewalk or a crosswalk (Fig. 3.4) for example, it is important that the flow of the group's movement isn't disturbed simply by having similar paths in mind. Perception analysis is done on a server external to the simulation.

While present in the simulation, traffic signs and lights do not influence much of pedestrian behavior. Agents do not wait in traffic lights, as should perhaps have been expected. This is because this perception had not been implemented at the time of development of this dissertation's project.

An urban environment of this scope should evidently feature cars that navigate the roads of the city. These cars should be able to reach a destination by moving through roads and respecting rules, such as signs and crosswalks. For this, an external simulation tool, ArchiSim [arc07], was employed. It allows for quick configuration of cars in a simulation of the same city environment as the Unity3D representation (Fig. 3.5). Thus, such cars in the external simulation tool can then be imported to the Unity environment using integration scripts (that were provided). The simulation is advantageous in the way that these cars respond to each other and attempt to their best not to collide with any obstacles, including other car agents. They respect road rules, accelerating in places where they can speed ahead, but also slowing down in intersections, stop signs, and if there is another car in front of it. This set of car agents was used in Unity by importing its own module to the three-dimensional environment.

The simulation is not without limitations, however. The way to configure the cars' initial positions and properties is through editing a text file descriptor. It is highly manual work and it is not evident how the values in the text file will translate to how they end up in the simulation. Moreover, in regards to their speed, they are usually fast (this is configurable) as long as the simulation is running with a high number of frames per second. This meant that slowdowns in the Unity3D simulation implied a slowdown in their speed. This slowdown proved to be quite drastic, leading to the feeling that they were extremely slow in comparison to pedestrians. Nevertheless,

Figure 3.3: A close-up of a pedestrian

Figure 3.4: Pedestrians crossing the road in order to reach their goal.

they allow for a more dynamic environment.

Figure 3.5: Car agents navigating through the roads of the city environment.

In summary, the main elements that comprise the city are:

- **Roads** where cars can circulate according to rules.
- **Sidewalks** where pedestrians can walk freely.
- **Buildings**, walls, bridges, roundabouts, and other static objects that fill up the environment.
- Poles, garbage cans, and other collidable **clutter**.
- **Parked cars** and vans lining the sidewalks.
- Moving **pedestrians** walking around according to rules.
- Moving **cars** driving around according to rules.
- Stop and yield **signs**.
- Pedestrian and vehicle **traffic lights**.
- **Crosswalks** in intersections.
- **Terrain** that fills up the background of the city.

The scope of the research implied the study of driver behaviors. As discussed in the previous section, such study using only agents in a simulated environment is disadvantageous in the way that scenarios are merely as good as the agents' rule sets. Thus, it was needed that humans could interact with the pedestrians inside the environment so as to extract the best driver judgment from them, and not from simple car agents. This required the use of a virtual cockpit, where the human driver could be immersed in and could drive around in. Thus, a virtual cockpit module was

employed for this purpose. The module consists of a simple family car where the driver takes on the driver seat and can see the environment ahead (Fig. 3.6). The car contains all the normal elements for driving, ie. a steering wheel, side mirrors, and a big transparent windshield. Also, a virtual GPS providing a top-down view of the car aims to help drivers orient themselves around the environment.

Driving in this virtual cockpit aims at realism. A gear box system allows users to drive like they normally would, contrary to many other driving simulations and games where the car is merely automatic in this regard. The car controls are quite sensitive and respond well to changes in steering, acceleration and braking. These parameters are also able to be changed at any time.

Figure 3.6: A view of the inside of the virtual cockpit.

Reaching a high feeling of realism in the driving experience necessarily requires a high level of immersion. For this, a virtual cockpit should not be controlled by a keyboard, since this would take away from the normal body movements and reaction times in real-life scenarios. Thus, two tools were used in order to increase immersion: a VR headset that displays the inside of the cockpit to users, and a steering wheel and pedal set connected to the controls of the cockpit.

Besides increasing immersion, the VR headset also facilitates visualization inside the cockpit. Without it, controlling the car and moving the camera around could not be done simultaneously. The reason for this is that by using a keyboard the car is controlled using one hand on the arrow keys for steering and the other for accelerating, braking and changing gears. The camera rotates through the use of the mouse. Thus, all this could not be done at the same time. This factor would change the driving experience and would not make it faithful to a life-like one, since in the latter one has full freedom over head movement for sight and simultaneous control of the steering wheel and pedals with hands and feet.

Nevertheless, the VR headset comes not without limitations. Many users are not accustomed to VR and their movement can be different from that in real-life. Users that are not familiarized with it tend to move their head around much more often and enthusiastically, and are much more

reactive to newer elements in sight. This can compromise the realism of data gathered through its use. Moreover, many users also experience symptoms of VR sickness. Feeling inside of a virtual environment but also removed from some degrees of freedom in regards to body movement, as well as fast movement and a frame rate different from that of the human eye can make subjects uneasy, uncomfortable, or nauseous. Naturally, this can compromise results in experiments done with the headset.

The HTC VIVE (Fig. 3.7) was able to be integrated with the Unity3D environment using the Steam VR tools. This allowed for easy switching between using VR or not, inside Unity. That ease of switching was especially useful during development.

The VRToolKit[1] was used in the project to act like an abstraction layer for VR and pedal controls. It was chosen so that development could be done even when not having the headset available for use. The tool kit allows the user to load which setup they wish to run the environment in (real VR or simulation).

Figure 3.7: The HTC VIVE headset (left) and Logitech G27 Racing Wheel (right).

In regards to controlling the car movement, a Logitech G27 Racing Wheel and pedal set was chosen (Fig. 3.7). The steering wheel feels very realistic, since it directly translates how a real car would be controlled. The wheel restricts fast movement, while allowing big turns at the same time. The pedal kit is very similar to that of a real car, and permits users to regulate car speed using a part of their body that would otherwise not be used were a keyboard chosen.

One difference in the control set is that the gear shifts are performed differently to that of a real car. Gear shifts are triggered using two handles on each side of the steering wheel. This makes the movement of switching gears feel very different. However, it was decided that this disadvantage did not play as heavy a role as the advantages of using a steering wheel and pedals.

3.2 Architecture

The environment for development, as mentioned, is highly modular. Scenes can be switched on and off according to the user's needs, using Unity's multi-scene loading. Each scene works

[1] available at https://github.com/ExtendRealityLtd/VRTK

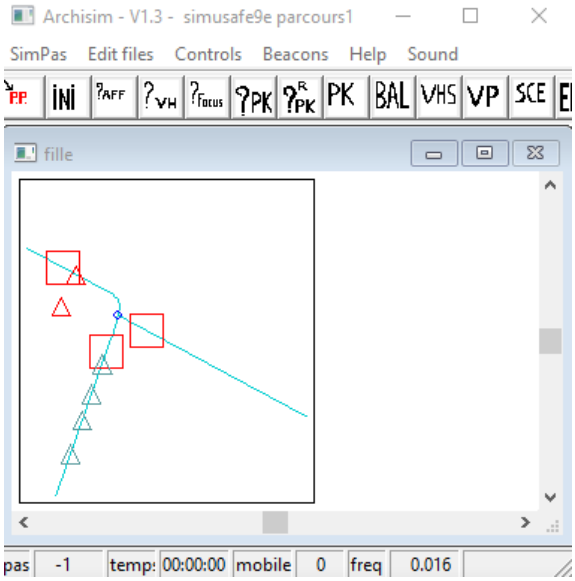


Figure 3.8: Archisim's simulation window.

as an additive for the base scene. Thus, they can be seen as modules, in a conceptual point of view. Since each scene adds different information and has its own purpose, they work as modular additives for the base environment. For studying driver-pedestrian interactions it was necessary that the car and pedestrian scenes were switched on. Their use implied the use of external tools: a perception server capable of processing each pedestrian's perceptions and an ArchiSim simulation that controlled the cars in the simulated environment (Fig. 3.9).

The perception server calculates perceptions in every frame, for every pedestrian. This perception processing is especially useful for collisions, since pedestrians should not collide with obstacles and with other pedestrians. Thus, this server was an integral part of the overall architecture of the system.

The ArchiSim simulation (Fig. 3.8) controlled the cars inside Unity environment. It runs an external simulation, and calculates each car's trajectory and behavior each tick of the run. This external simulation featured a layout that is similar to the one in Unity. It is also highly configurable, and it is possible to add or remove cars in any part of the city at will before beginning the simulation. Their speed and initial direction is also configurable. Thus, since cars are an integral part of the city environment this simulation was also crucial to the study at hand.

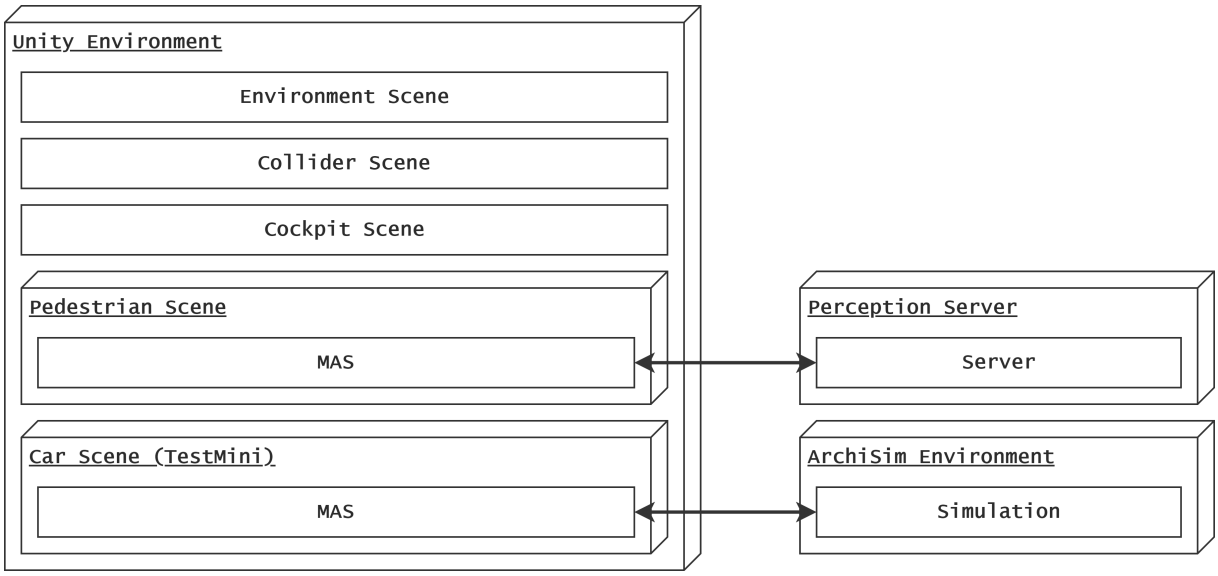


Figure 3.9: A diagram of the architecture of the environment.

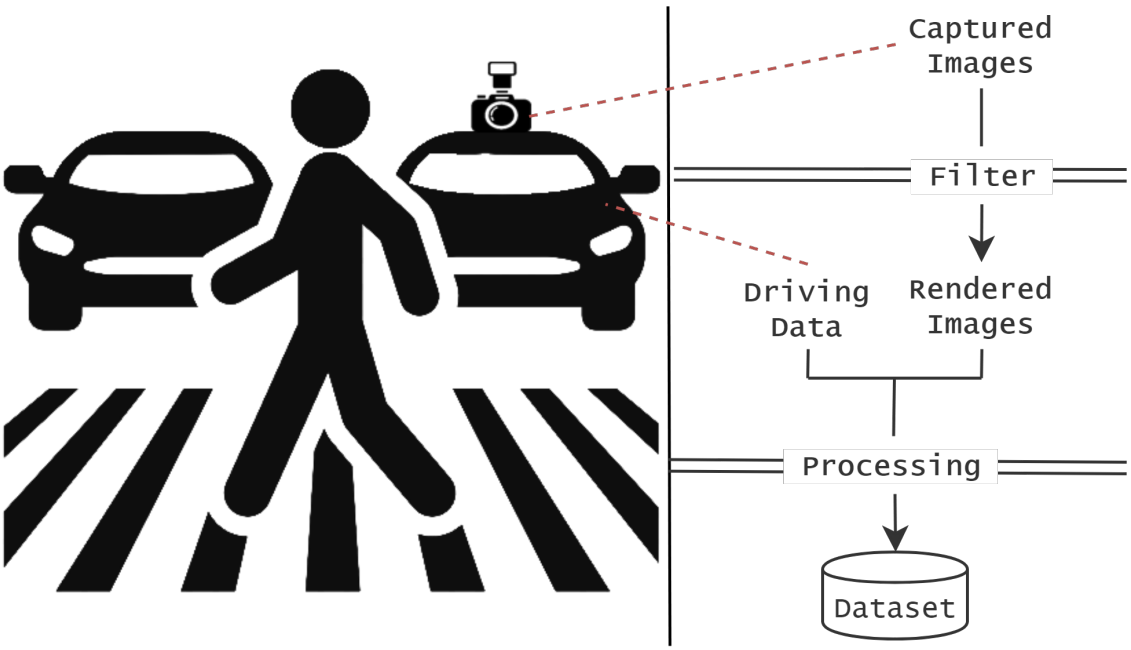


Figure 3.10: A diagram of the pipeline for collecting data.

To obtain data about driver-pedestrians, a pipeline architecture (Fig. 3.10) handles collecting data from the Unity architecture and the outputting it in a concise dataset.

During driving, information about the car and driver is collected. This information can provide insight into driver actions and reactions in pedestrian encounters. It is relevant that data about steering, speed, acceleration and braking is collected. These actions are fundamental in encounters, because they are ones of the most crucial variables that ensure that encounters go smoothly

and accident-free. Drivers can steer the car away from pedestrians and obstacles in encounters, as well as braking or keeping speed low during them. On the other hand, if drivers are not cautious or experienced they may choose to accelerate and take dangerous actions. Thus, it is important that data about these actions is gathered. Even more, it is important to analyze where a driver is looking during driving. If they are not looking ahead they may not see incoming obstacles. As such, it is important that such data is kept as well.

Besides driver information, a camera instrumented into the car collects visual data during driving. This camera can provide insight into the visual cues and elements seen by the driver. Raw camera data does not provide much knowledge on its own. Thus, it is necessary that it is processed. This processing ensures the creation of segmented images, the calculation of pedestrian groups, and other useful information. This part of the methodology is explained in the next section.

The gathered car data is coupled with the visual information for the creation of pedestrian action maps and driver maps. These are stored into a dataset that can be the basis of a predictive model for intention inference.

This pipeline acts as a black box that takes as input raw visual data and instrumented car data and outputs tables and maps of occurrences during driving.

The study of driver-pedestrian interactions required the gathering of such driving data and visual cues that are given to the driver. Thus, this necessity meant the construction of the pipeline that was described in this section.

3.3 Extracting Information in Driving Scenarios

Driving experiences are activities that provide stimuli of various kinds to drivers. Competent drivers will acknowledge and react appropriately to the constant stream of stimuli that they receive while driving. Stimuli may be of visual, auditory or contextual kinds, among others. The first and latter ones are the focuses of this body of work.

Visual cues are crucial to driving. Drivers should react to visible obstacles, as well as any non-relevant visual information in sight. Should the driver see a pedestrian crossing the road in front of him, he should be able to process this presence in sight and know that a pedestrian is present. Cars, obstacles, signs, and many others also constitute visual information that an experienced driver will assimilate without much conscious thought. Besides, obstacles can be only partially in sight, through obstruction in visibility by other obstacles. Ideally, a pedestrian that just appeared in sight and a pedestrian that has remained in sight for long should both be regarded as equally relevant presences in the field of view.

The assimilation of visual information does not mean that drivers should react similarly to the same cues. As seen in section 2, drivers' profiles are not one-dimensional, and so their reactions are not all similar. While upon seeing a pedestrian cross the road a driver may choose to stop, others may simply try to keep a constant speed, swerve away from it or even speed up, for example.

Thus, taking visual cues in consideration in driving is crucial to investigate how drivers read pedestrians' intentions and how they manage their own.

Image segmentation techniques are interesting tools to be employed in the field of image analysis. Their uses are multiple, ranging from organ recognition in the medical field [MNA16][PXP00] to complex recognition tasks. This flexibility makes such techniques an area of interest for researchers. Such techniques allow for an image to be classified in its components, thus allowing for extracting semantic knowledge from pixels, or for identifying different instances of objects as clusters of pixels in the image. Semantic segmentation aims at attributing meaning to groups of pixels that share a set of characteristics. It allows to identify elements of the same kind in an image. Instance segmentation aims at isolating each cluster of pixels that represents an individual object in an image, without immediately attributing it some meaning. Techniques for both semantic and instance segmentation are quite varied.

3.3.1 Semantic Segmentation

If one aims at attributing a certain meaning or category to pixels in an image, semantic segmentation allows for this. Pixels in the image to be classified may fall into one or more categories. This allows for the creation of areas of interest in which pixels clustered together form a bigger, more important and defined role than their basis definition of color.

This tool is relevant in extracting information from driver-pedestrian interactions. Drivers that pay attention to road are attributing semantic meaning to each element present in the field of view. By replicating this using semantic segmentation and categorizing resulting segments, one can extract information about the presence of different elements in the scene. For example, in an image containing a crosswalk and pedestrians waiting to cross, a segmented image could provide information about which pixels define the pedestrians and which ones define the crosswalk.

From knowing the present types elements in the field of view, one can relate it to the drivers' immediate reaction. For instance, if associating images and a driver's reaction an image contains pixels that correspond with pedestrians this may explain why the driver's reaction in sequence was to brake. The same applies to other elements, be them other cars, traffic signs or obstacles present in the road. Thus, knowing elements that are present throughout the driving experience can allow to infer over correlation of their presence (or absence) and the actions taken during driving.

In real driving scenarios there can be an overwhelming quantity of elements that appear and disappear of the driver's view and so it is important to define which ones play the biggest role in driving. Not including all of them, but some relevant visual elements that influence driving and should be categorized may be, in no particular order:

- Pedestrians, whether crossing the road, near the curb, on the sidewalk, etc.
- Cars, whether moving or parked.
- Obstacles in sight, be them lamp posts, garbage cans, etc.
- Cyclists and motorcyclists.
- Buildings and infrastructure.
- Animals.
- Traffic lights, for pedestrians or cars.

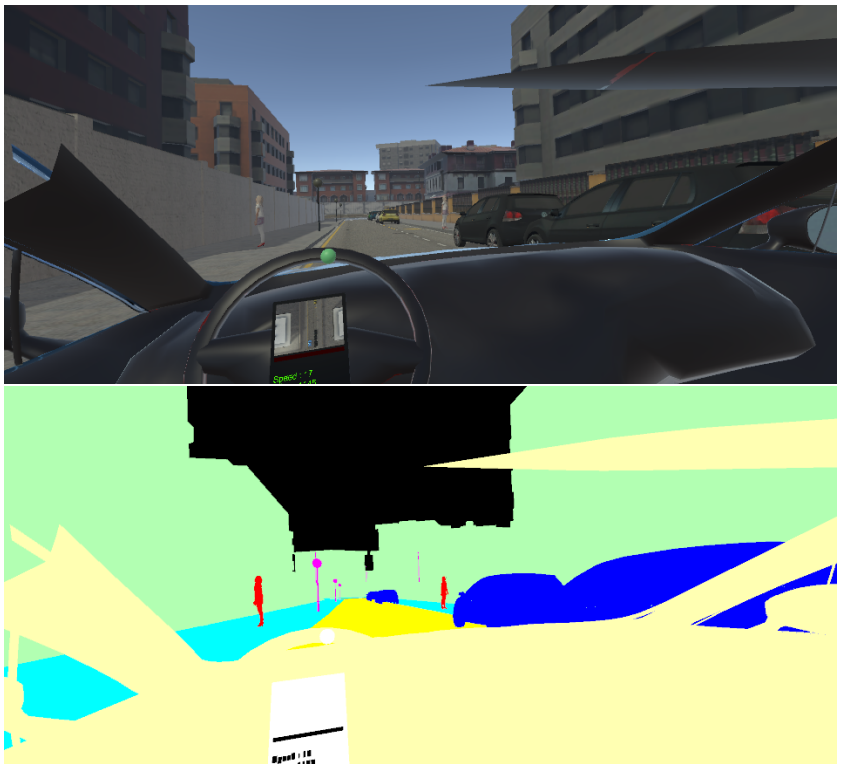


Figure 3.11: An image before segmentation (above) and after (below).

- Traffic signs.
- Crosswalks and sidewalks.
- Indications on roads.

Knowing this, image segmentation is performed over the driving experience in the environment. Images are decomposed into these main elements and the resulting image contains only a set of the colors that are attributed to them. The advantage is that this new image contains clusters of colors that are now related to one another, in contrast to the original image where each neighboring pixels' colors would easily be different and not immediately related. Such result allows for the extraction of knowledge of the driving inside of the environment, for analysis. It was decided that the elements described above would be the main elements to identify in a scene.

The result of a segmented image is a colored image in which pixels of the same color share some characteristic. For example, in Fig. 3.11, pixels that are red belong to the category "Pedestrian" while pixels that are lime green belong to "Building".

As the development environment was Unity3D, one can make use of its camera and shader pipeline to apply image segmentation. For this, a library[2] that makes use of renders to encode elements in the images was employed as an off-the-shelf solution to image segmentation. The driver's cockpit is supplied with a camera in order to record a sequence of images during driving. This camera is equipped with other hidden cameras that bypass Unity's normal rendering system

[2] available at https://bitbucket.org/Unity-Technologies/ml-imagesynthesis

and instead equip shaders to obtain a new rendered output. As for image segmentation, the shader is able to perform it if every different category of element is listed as a different object layer in Unity. Thus, Unity objects were laid out in these layers so that the shader would see them as part of the same element type if they belonged to the same layer. For example, all buildings were inserted in the "Building" layer. The camera encodes a different color for each layer. This color does not change between runs and could be configurable. For the building layer, the resulting color was a light green.

Had the development environment not been inside a virtual 3D world, another method of segmentation would have had to be employed. Among many different techniques, Mask-RCNN techniques similar to those mentioned in section 2 are by far the best performing. They allow for semantic segmentation similar to that obtained by the aforementioned library, with the advantage of the possibility of its use in any image, virtual or not. However, they usually require large training and testing sets, and training times to use.

It was investigated whether the employment of a Mask-RCNN approach to segmentation could be used in Unity3D. Most recent approaches to Mask-RCNN make use of the TensorFlow tool to facilitate training and testing, as well as performing them more efficiently. Integration between TensorFlow and Unity was, as investigation implied, not easy. Many TensorFlow approaches are based on Linux operating systems as well, which clashes with the used version of Unity3D being Windows-only. Moreover, Mask-RCNNs rely on training using large datasets, which were not present for this specific Unity environment and had to be created. For these reasons, this approach was discarded and the library was employed.

The resulting image conveys information about the present elements in the scene. There are as many different elements present in the scene as the number of colors in the result. From here, one can extract a simple table of presences for each element that is able to be identified in the scene. Elements are tagged with a truth value for each resulting image. If an elements corresponding color is present in the image it is tagged with *true*. If not, with *false*.

Algorithm 1 Element Isolation in Segmented Image

```
1: segmentedImg ← segmentImage(path)
2: elementColor ← color(element)
3: table(element) ← false
4: for row ∈ segmentedImg do
5:     for pixel ∈ row do
6:         if color(pixel) = elementColor then
7:             table(element) ← true
8:         else
9:             color(pixel) ← BLACK
```

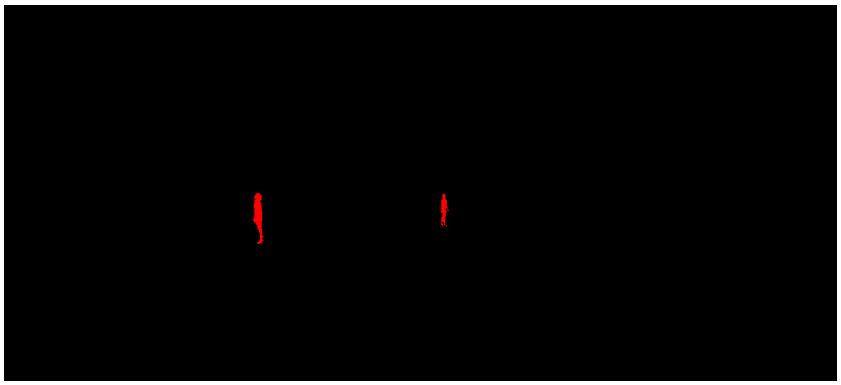


Figure 3.12: A mask applied to the segmented image.

To create such a table, one needs to analyze the image for each element and check if at least one pixel is of its corresponding color. For this, an algorithm that runs comparisons between pixels and the elements' colors is deployed (Algorithm 1). If the pixel color and element color are the same, then the pixel is kept. If not, the pixel is turned black (Fig. 3.12). This mask is created for each element color. The result makes present elements salient. If there is at least one pixel that is not black, then that element is present in the image.

After running a mask for each element type, the table is created. This table is created and saved for each image and it allows to quickly check which elements were present at any time without having to run the image segmentation again. For the example in Fig. 3.11, it would have the configuration of Table 3.2:

Table 3.1: Presence table for Fig. 3.11

Element	**Present**	**Color**
Pedestrian (sidewalk)	*true*	RGB(255, 0, 0)
Pedestrian (crossing)	*false*	RGB(255, 180, 180)
Pedestrian (near crosswalk)	*false*	RGB(90, 0, 130)
Car	*true*	RGB(0, 0, 255)
Obstacle	*true*	RGB(255, 0, 255)
Building	*true*	RGB(180, 255, 180)
Sign	*false*	RGB(180, 0, 0)
Traffic Light	*false*	RGB(0, 180, 0)
Sidewalk	*true*	RGB(0, 255, 255)
Road	*true*	RGB(255, 255, 0)
Crosswalk	*false*	RGB(180, 180, 180)
Cockpit	*true*	RGB(255, 255, 180)

Interactions differ with regards to the pedestrian's position when sighted. A pedestrian that is seen crossing the road should be considered as different from one that is waiting to cross, or one that is simply walking on the sidewalk. If a pedestrian is seen crossing, the driver is able

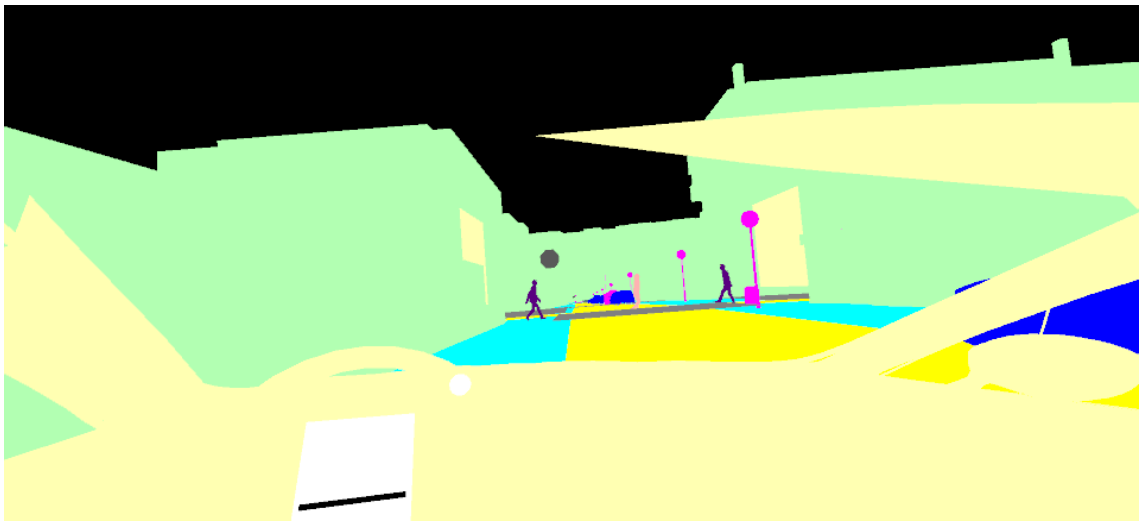


Figure 3.13: Pedestrians near the crosswalk (purple) and a pedestrian crossing the street (beige).

to quickly infer over their intentions: they wish to cross the road. However, if the pedestrian is seen waiting on the edge of the crosswalk, a driver may not as easily infer that they want to cross. The pedestrian may wish to cross, but it is uncertain to determine. Hence, pedestrians' position in relation to other elements should be considered when analyzing an interaction.

For this reason, pedestrians were sub-categorized so that their positioning plays a role for later analysis. Pedestrians fall into three categories:

- Pedestrian is crossing the road.
- Pedestrian is standing near the crosswalk.
- Pedestrian is walking on the sidewalk.

This sub-categorization was done through use of Unity3D's layer system and how image segmentation is being performed. When approaching crosswalks, pedestrians trigger colliders that change their layer to "Near Crossing". Likewise, when they are crossing, the layer changes to "Crosswalk". When they get back on the sidewalk, the layer is reset to the default pedestrian one (Fig. 3.13).

During driving, frames should be captured so as to store what was observed at all times by the driver. Such frames are merely images comprising the field of view of the driver's eyes, including any elements present on the road, above it, or inside the car itself. These frames are the basis of the research using visual cues. If all possible visual information is stored, every occurring visual cue can be explored and analyzed. Thus, for every recorded frame, semantic segmentation is performed and a table of presences is stored. The table of presences provides a part of the needed data to be analyzed in posterity. It ensures that less storage is used in gathering information, since its size is much smaller than that of the images stored during driving.

The use of semantic segmentation serves the purpose of not only associating images with the elements seen within them, but also allows for such data to then be paired with other driving data for analysis and correlation.

Table data should also be coupled with the time of recording, as well as the drivers' actions in terms of vehicle control at that time.

One can assume that the presence of pedestrians in sight will influence driving there on forth. Thus, since the truth values of the tables imply their presence in the image, these tables and coupled data will provide the basis of knowledge for relating possible pedestrian sightings with drivers' actions in time intervals before and after the table has been stored.

This assumption withstanding, not all available information in an image can be extracted through the use of semantic segmentation. Pedestrian sightings are not all equal and independent events. One limitation of semantic segmentation for this purpose is that it does not take into consideration the effects of the presence of more than one pedestrian on the sidewalk. When someone is willing to cross, it may not be the only one to want to do so. Thus, a group of people may be waiting near the crosswalk to get across.

The existence of pedestrians in different sides of the road is also a very different event than that of only one of them wanting to cross. The presence of groups, either on the same side of the crosswalk or on opposite ones should also be taken into consideration when analyzing driver-pedestrian scenarios. For this reason, it is important that when analyzing a recorded image of a driving experience, the number of people in sight be considered. As aforementioned, semantic segmentation only attributes meaning to groups of pixels that are seen as elements of the same category. Since pedestrians should all belong in the same main category, the technique provides no information about the amount of in sight. For this reason, the technique needed to be coupled with another segmentation technique, so as to surpass such limitation.

3.3.2 Instance Segmentation and Pedestrian Groups

The influence of the number of elements on the scene could not be solved through the sole use of semantic segmentation. There arose a need to evaluate the quantity of each relevant type of element on the scene. The number of parked cars, the number of moving cars, the number of pedestrians crossing the road, as well as many others provide useful information that would help understand driver-pedestrian scenarios.

As an urban environment is crowded, there is not only a single instance of each element on the scene at all times. Thus, such instances needed to be identified in the scene as well. Instance segmentation came as the solution to this problem. This type of segmentation isolates each unique object on the scene as its own cluster of pixels. Two pedestrians alongside each other would not be treated as one pixel cluster, but instead as two.

Separating objects on the scene is a famous computer vision problem, and many approaches are available for solving it depending on the desired outcome. Some recent approaches rely on training and deploying deep networks to process the image and output pixel clusters or edges of objects. As discussed previously, the deployment of a deep network in the development environment proved to be quite arduous and as such was discarded. It was decided that the same library that performed semantic segmentation to be the one to provide instance segmentation as well.

Inside Unity3D, each object present in a scene is treated as an instance of a type of object. Each instance is tagged with an unique identifier. Thus, the approach to instancing each one in a rendered image is quite parallel to the instancing of each type of element. Instead of tagging each type of element as its own layer, each object's identifier is encoded as a unique color and that color is displayed on the rendered image. As such pedestrian clusters appear as more than one color.

The methodology employed for extracting elements present on the scene is useful for this next step. That resulting image is used as a mask that is applied over the new segmented image. For every pixel that is black, that pixel and its position is ignored. If it is not black, it is kept. The image obtained after using the mask has the same shape and edges of the mask, but may have one or more colors. The presence of more than one color means that more than one element of that type is present on the scene.

Algorithm 2 Instance Isolation in Segmented Image

```
1:  instanceSegmentedImg ← segmentImageInstance(path)
2:  elementColor ← color(element)
3:  tableCount(element) ← 0
4:  counted ← empty list
5:  for row ∈ mask do
6:      for pixel ∈ row do
7:          if color(pixel) = elementColor & color(instanceSegmentedImg(pixel)) ∉ counted then
8:              tableCount(element) ← tableCount(element) + 1
9:              counted(color(instanceSegmentedImg(pixel)).insert(color(instanceSegmentedImg(pixel)
10:         else
11:             color(instanceSegmentedImg(pixel) ← BLACK
```

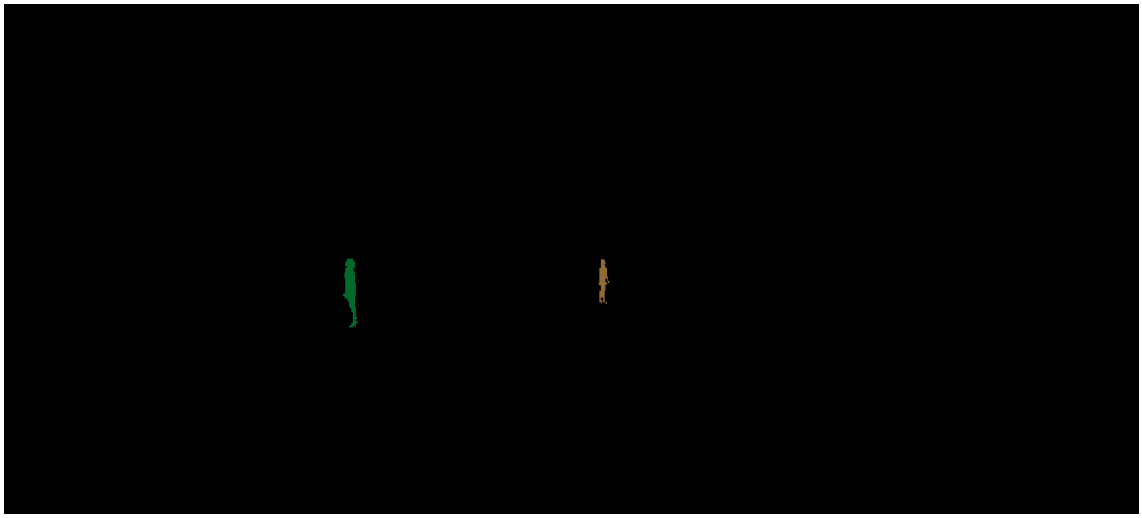


Figure 3.14: Different instances of pedestrians.

The algorithm (Alg. 2) creates a new colored image (Fig. 3.14) but also creates a table similar to the table of presences of elements. The new table simply provides the value of how many different colors were present. Since the number of different colors translates the number of objects

of that element that were spotted in the image, that number means the amount of entities of that type on the scene.

Table 3.2: Instance counting table for Fig. 3.11

Element	Number	Color
Pedestrian (sidewalk)	2	RGB(255, 0, 0)
Pedestrian (crossing)	0	RGB(255, 180, 180)
Pedestrian (near crosswalk)	0	RGB(90, 0, 130)
Car	4	RGB(0, 0, 255)
Obstacle	3	RGB(255, 0, 255)
Building	11	RGB(180, 255, 180)
Sign	0	RGB(180, 0, 0)
Traffic Light	0	RGB(0, 180, 0)
Sidewalk	2	RGB(0, 255, 255)
Road	1	RGB(255, 255, 0)
Crosswalk	0	RGB(180, 180, 180)
Cockpit	1	RGB(255, 255, 180)

Simply counting the amount of pedestrians on screen does not provide information about their position or if they are grouped up with other pedestrians. If two pedestrians are on two different sides of the road the algorithm counts two pedestrians, but there is no way of extracting their group status from this number alone. Thus, some other image operations will have to be performed in order to make up for this limitation.

What can be considered a pedestrian group is quite ambiguous. Pedestrians may be side by side in an image at a given time, but this does not immediately imply that both pedestrians are willing to cross. Moreover, two sets of pixels that are side by side in the image may not mean that the pedestrians are even side by side. Perspective can make two points that are quite distanced on the 3D world appear like they are quite close on the 2D plane. Therefore, distance also needs to be factored if one wants to correctly analyze pedestrian groups. The three things that should be considered when checking for pedestrian groups should be:

- Their **closeness** in the image: pedestrians should be close in order to be qualified as being in the same group
- Their **distance to the camera**: pedestrians that are close in the image do not mean they are close in the real world. Distance should be factored as well
- Their **direction**: pedestrians may be close but if they are facing different ways they convey different intentions.

Gathering information about the two last points will be the focus of the next section. Assuming that two pedestrians are side by side in the world, one can check if they are close by analyzing the

pixels in an area of their instance's pixels and checking if another pedestrian's pixels make part of such area.

For every pedestrian, the algorithm takes the instance segmented image and checks the pixels of pixel clusters' edges for a possible other pedestrian in the neighborhood. If a color that is not that pedestrian's instance color is present in such area, that pedestrian is close to the one being checked. This process goes on to the next pedestrian detected and so forth. If the next pedestrian is checked as close, the program runs another check on it, and if it detects more it goes on to the new one. The number of times this process is repeated is the number of pedestrians that are close to the first one. For each pedestrian that is detected a value depicting group size is stored. Only clusters' edges' pixels are check so as not to render this check as intensive for processing. If the algorithm goes back to the first and and there are still pedestrians that have not been checked, this means they are not close to the first one and another recursive check is run on it.

This algorithm also takes into account the distance of the pedestrian that is detected as close on a given check. If this pedestrian's distance is very different from that of the first one, it is skipped.

Algorithm 3 Instance Isolation in Segmented Image

```
1:  procedure SET GROUP SIZE(pedestrians)
2:      for pedestrian ∈ pedestrians do
3:          pedestrian.checked ← false
4:          detectedPedestrians ← empty list
5:          pedestrianPixels ← getPedestrianPixels(pedestrian)
6:          distance ← getDistance(pedestrian)
7:          for pixel ∈ pedestrianPixels do
8:              if surroundedBySameColor(pixel) ≠ true then
9:                  detected.insert(check(distance, detected, pixel))
10:         pedestrian.checked ← true
11:         pedestrian.groupSize ← countUnique(detected)
```

Algorithm 4 Check Neighborhood for different colors

```
1:  function CHECK(distance, detected, pixel)
2:      for npixel ∈ neighborhood(pixel) do
3:          if color(npixel) = BLACK then return null
4:          else
5:              if color(npixel) ∉ detected & getDistance(npixel) = distance then
6:                  detected.insert(color(npixel)
7:                  check(distance, detected, npixel)
```

Group size is stored for each pedestrian. This value will be used for analysis on the influence of groups on driver judgment. Demographics were not considered in this scenario, since the pedestrians in the simulation are almost homogeneous (there are only two different models).

As discussed before, group size calculation has to take in consideration the direction of pedestrians in such group as well as their distance between one another. Thus, it is important to evaluate such measurements for each pedestrian at all times.

3.3.3 Element Distance and Direction

It comes as evident that at all times, drivers are evaluating not only the visibility of elements in sight but also their position in relation to the car. Obstacles that are getting closer are seen as threats and must be avoided. This reaction depends not only on the driver itself but also on the predictability of the obstacle's movement. Indecisive pedestrians or pedestrians that jaywalk may be much harder to dodge than a simple pole. Thus, drivers rely on their judging of distances, speeds and times to collision to all elements and what they represent as threats to driving. Naturally, in accident-free driving scenarios, this evaluation can be mostly considered as successful.

In this context, obtaining information from an image about each element's distance could be done by using a simple depth map of the scene. Depth maps provide information about distance by setting a pixel's color to a darker one if the object that is seen through that pixel is closer to the camera. Thus, this creates a grayscale image in which distanced objects are light and closer objects are dark (Fig. 3.15).

Unity's render textures provide all information needed to create depth maps. Such texture was applied to a camera which results in the final image being the desired depth map. The library used for segmentation also provided the tools to do this easily.

Pedestrian distance is given by the closest of all pedestrian's pixels on the image. This ensures that a pedestrian's distance is based on the nearest possible distance that it is possibly able to be hit by the car.

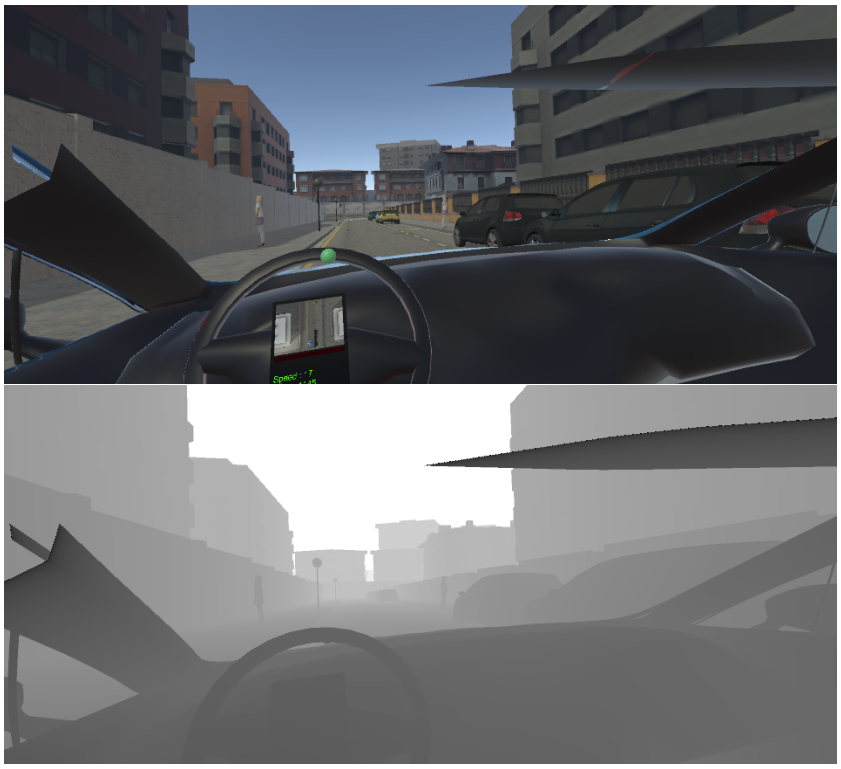


Figure 3.15: Depth map view.

For dynamic elements in the scene such as cars and pedestrians, it is important that where they are facing is accounted for. Pedestrians that are facing away from the crosswalk may give the impression that they do not want to cross, and vice-versa. In a group of pedestrians near the crosswalk, it is much easier to infer that the ones facing the other side of the crosswalk are meaning to cross, and harder for ones facing away from it. Moreover, pedestrians tend to look to either side of the road when crossing. Of course, not always does this happen. If a pedestrian is jaywalking it might have not looked to see if the road is clear. Pedestrians that are fully confident on driver yielding when approaching a crosswalk may also not look at other places than the other side of the crosswalk.

The same applies to cars. Cars that are facing the same way as the driver are seen as going forward (if not parked). Thus, collaboration between them should be employed so that the one behind does not hit the one in front, for example. Cars facing the opposite way or facing a different way are seen as obstacles much more easily, and road rules allow for inference of yielding in such scenarios.

Thus, not only distance is relevant for inference but also direction. Given this, calculating direction in an image is not trivial. Elements are not simple planes and are often complex shapes which makes guessing their direction harder for drivers. Moreover, obstruction of visibility may not allow for such guessing.

Normal maps are used in various contexts much like depth maps. They provide the direction of the element of each pixel on the scene (Fig. 3.16). As the world is three dimensional, all three axis have to be considered. The RGB color space allows for this consideration easily, by applying a different intensity of each color in proportion to the value of normal vectors of each element. Normal vectors are simple three dimensional vectors that provide information over direction in the way that they are perpendicular (thus, normal) to the object's surface. In this context, the RGB color space outputs a value such that that value provides information over the normal vector calculated there. For each pixel the value is calculated using each axis' value over the norm of the normal vector and multiplied by the max RGB value, 255 (Eq. 3.1). This provides a RGB vector wherein the values are all between 0 and 255 and a color corresponding to a direction.

$$RGB = (255\frac{n_x}{|\vec{n}|}, 255\frac{n_y}{|\vec{n}|}, 255\frac{n_z}{|\vec{n}|}) \tag{3.1}$$

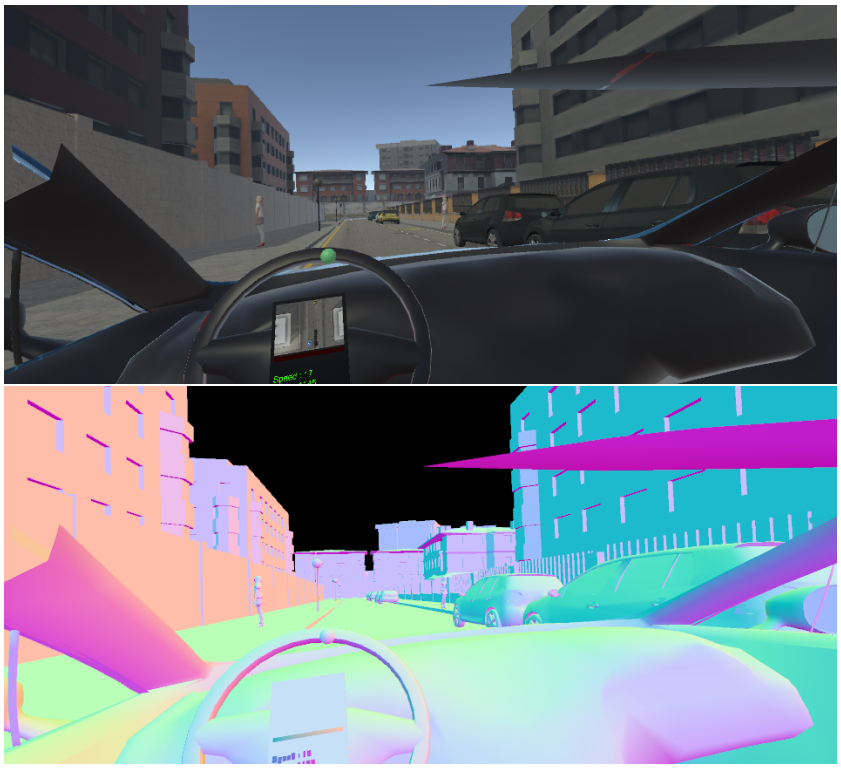


Figure 3.16: Normal map view.

Obtaining the direction of the pedestrian is, as mentioned, important for the study of the relation between their body language and driver inference. Of course, direction alone does not provide all the information needed to get a complete report of their communication. Among the most important factors in body language, one of them is the establishment of eye contact between the driver and the pedestrian. Pedestrians that look at a car when it is approaching them are evidently aware of their presence, in a way that other cues do not directly allow for. Thus, pedestrians that gaze to the direction of the car are an important element to be factored in this study. For this reason, the normal map alone would not provide full information. The normal map would give us the direction of their body in relation to the car, but no direct information about the direction of their gaze.

To better understand pedestrians' movement one could use several techniques. Captured image data could be analyzed for optical flow of the elements seen. This optical flow allows for an estimate of the future movement of such elements, and thus distinguish static from dynamic elements, and the latter ones by their apparent speed. Pedestrians' movement and gazes could also be estimated through the use of face detection algorithms or pose estimation networks. Coupling all these could provide a better outlook at a pedestrian's current and future relative positioning, and an insight into their body language and stance. However, these are quite heavy in performance and were left out in this study.

For this reason in this study pedestrian models were combined with an identifier on their model's face that would allow for checking their gaze. This identifier moves with the pedestrian

movement as if they were one. When performing image segmentation, this identifier is accounted for. If a pedestrian is spotted in segmentation, but their gaze identifier is not, that must mean that he is facing away from the driver. Moreover, if it is only partially visible this yields incomplete information about their gaze. If in segmented images their gaze is spotted this does not mean they are facing the driver. For this reason, the segmented image is coupled with the obtained normal map to check the direction of their gaze (Fig. 3.17).

If a pedestrian's gaze is facing a certain threshold that corresponds to a similar or parallel vector to the vector of the difference between their position and the driver's position, then it is gazing at the car. This approach was used for analysis over if pedestrians looking at the car could influence driver inference over their behavior.

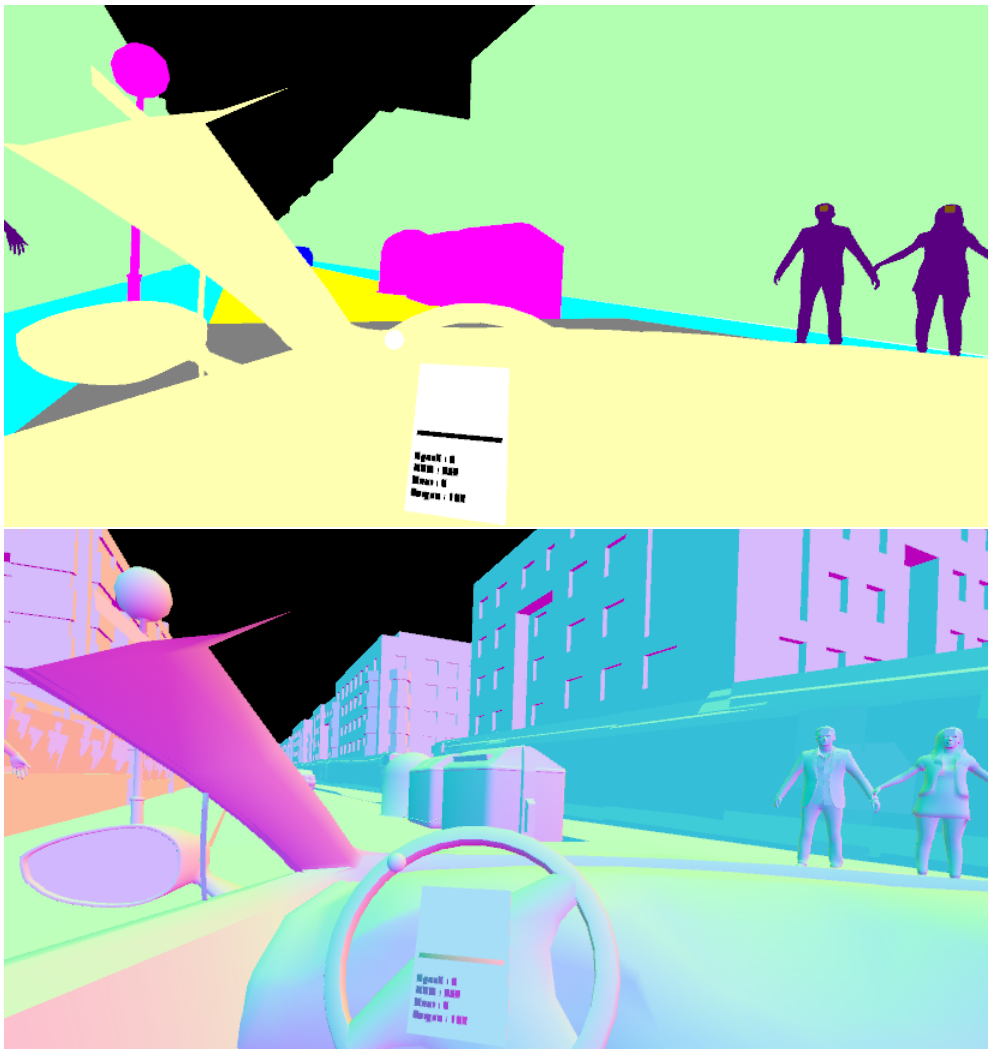


Figure 3.17: Pedestrian gaze as the brown plane in segmentation.

In regards to other types of body language, a person's pose could give some insight over their intentions. Pedestrians that raise their hand to signal they want the car to stop provide direct information over their intention to cross. If body language is fast-paced, that may imply they are

in a hurry to cross. On the other hand, pedestrians that look relaxed and make little gestures will be seen as lenient and that they are willing to wait to cross. All this should be considered for a complete study of pedestrian intentions.

As mentioned in chapter 2, approaches that aim at predicting pedestrian crossing using body skeleton models exist. These approaches apply a model over the pedestrian's pixels that defines each limb's position and orientation. They do not add new information on hand direction, gaze direction, etc. Nevertheless, such approaches could be coupled with the information gathered with groups and gaze analysis to obtain better information over their intentions. Such was the final aim of this methodology. However, implementation with such approaches relies on the use of complex deep networks and is therefore very intensive in processing. Coupling this with Unity and the segmentation systems implemented would make the system suffer greatly in performance. This was a big limitation, since VR demands high performance so as not to be uncomfortable and to users as well as assuring their immersion. Using such methodologies in Unity is also quite arduous since TensorFlow's integration with Unity is not trivial.

The used pedestrian models worked as skeleton models and did have different independent limbs and joints, which would be very useful for this approach. However, their model and locomotion system was not very diversified in terms of possible gestures and behaviors before crossing. It was noted that pedestrians showed little variation in movement before crossing. They would not signal the arriving cars and pedestrian, and there were no different possible stances for their regular posing. Of course, this poses as a limitation of the used environment and would render complete body analysis quite useless. Thus, the gesture analysis was discarded for the final project state and planned for future work.

3.4 Pedestrian Action

Studying pedestrian intentions cannot be done solely on single image analysis. This type of analysis leaves out an important method for humans to infer over pedestrians: contextual information. Movement, state changes (from crossing the road to simply walking on the sidewalk), pose changes, and many other factors are gradual changes that do not differ greatly between two sequential captured frames. Thus, it is important that information about pedestrians is stored and compared to previous information, and new information can be predicted. For this reason, it was necessary to create a system for studying a pedestrian's journey throughout the driving experience. This was named its history. From its history, one can point out any state changes, and this type of change is easily visible on a map.

This type of information visualization helps in understanding a driver's reactions if also mapped out alongside the pedestrian's history.

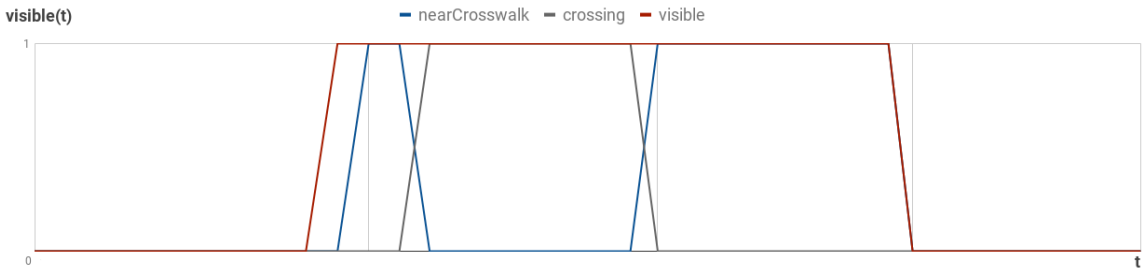


Figure 3.18: Pedestrian map state changes for a driving run.

The pedestrian map in Figure 3.18 shows the change of states for a pedestrian encounter. Pedestrian states are fixed on a scale of zero to one. Like boolean data, zero means the state is not active. Likewise, one means the state is active. The pedestrian is visible throughout a certain number of images (the red line), until it disappears from view. During this visibility window, it crosses the road. This can be seen from the appearance of two other states: *nearCrosswalk* and *crossing*. First, the pedestrian is walking on the sidewalk. Then it approaches the crosswalk, and crosses it, changing its state to *crossing*. Then it walks along and disappears from view eventually while still near the crosswalk.

This map is also paired with other maps depicting group size, gaze visibility and direction and distance to the car. Agglomerating these maps for every pedestrian provides data about ones that were encountered and under which conditions they were encountered.

Pedestrians that were encountered multiple times will have several different intervals of visibility. This ensures that it is not treated as a separate instance every time it may pop in and out of view due to obstruction.

Driver behavior is treated similarly to this. The only type of driver reaction that is not consisting of somehow changing the driving itself is the head movement that the person behind the wheel is performing. This head movement and gaze is vastly important for analyzing reaction and which objects tend to influence driving.

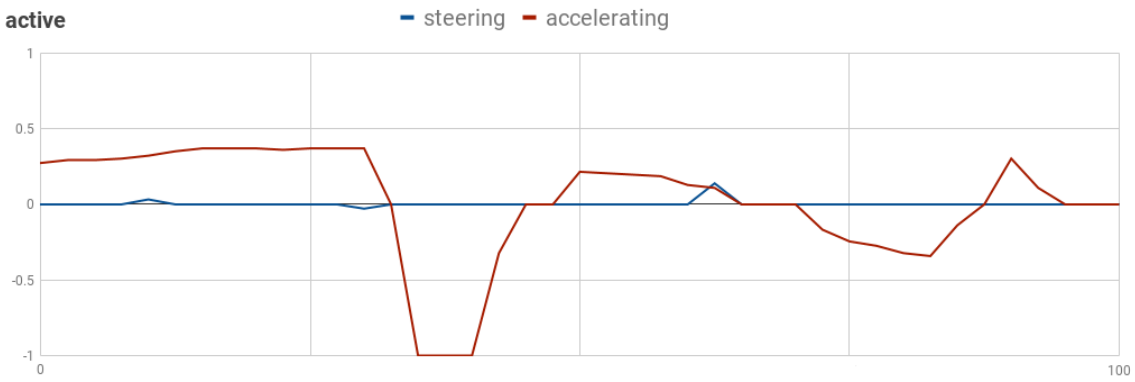


Figure 3.19: Driver map steering and accelerating changes for a driving run.

Figure 3.19 describes the sequence of driving actions over the same driving run as Fig. 3.18. Positive acceleration values imply the vehicle is accelerating. Otherwise, values that are negative imply the driver was braking. In terms of steering, positive values depict steering towards the right, whereas in areas where the values are negative the driver is steering left. When the pedestrian is first visible, the driver does not immediately react to it in driving. When it approaches the pedestrian it quickly brakes and waits for it to cross. Then, it accelerates when the pedestrian is almost done crossing and attempts to gently steer away from its direction. It then resumes driving as normal when the pedestrian is no longer a threat.

Combining all these maps provides a useful tool to tell the driving story through mapping. Therefore, this information, alongside visibility tables is what is needed for analysis over driving scenarios.

3.5 Data Collection and Preparation

As mentioned previously, image segmentation techniques and other algorithms process information over each driving image. A lot of information is obtained through them, and thus should be handled accordingly. To get a good look at driver pedestrian scenarios, and due to their complexity, all data that can be explored should so be.

Using instrumented vehicles during driving proves useful for such analysis. Dashboard cameras can provide all needed visual data for driving, and steering wheel and pedal sensors inform about exact metrics that could not otherwise be gathered.

3.5.1 Data Storing and Formatting

Evidently, there arises a need to organize such data. For this study several CSV (Comma-Separated Values) file templates were followed. These CSV files serve as spreadsheets where data can be quickly visualized and accessed. Three files were created for each run:

- A file that stores data over visible elements on each image, and their cardinality.
- A file that stores data about all driver metrics on a given time frame.
- A file depicting pedestrian history, and associated driver history.

Each file plays a role in gathering results from the study. The scope of each file funnels on a certain part of the driving experience that is to be analyzed. That is, the visual cues therein, the driver reactions at all times and the pedestrian encounters and their conditions.

The first file follows a similar structure to the visibility data mentioned previously, only transposed (Table 3.3).

Table 3.3: CSV template for image segmentation result files

frameNo	**t**	**El$_1$(presence)**	**El$_1$(count)**	**El$_2$(presence)**	**El$_2$(count)**	**...**
$\mathbb{N}_0$	$\mathbb{R}_{>0}$	*true / false*	$\mathbb{N}_0$	*true / false*	$\mathbb{N}_0$	**...**

The values in this table are organized as such:

- **frameNo**: an integer greater than or equal to zero that describes the position of an image in the sequence of images.
- **t**: a positive real number describing the time of the run in seconds in which the image identified by *frameNo* was taken.
- **El$_x$(presence)**: a boolean translating element x's presence in the scene.
- **El$_x$(count)**: an integer greater than or equal to zero translating element x's number of instances in the scene.

The second file contains the following structure (Table 3.4):

Table 3.4: CSV template for driver metrics.

frameNo	t	speed	steering	accelerating	braking	currentGear	headAngle
$\mathbb{N}_0$	$\mathbb{R}_{>0}$	$\mathbb{R}_{>0}$	$\mathbb{R} \in [-1,1]$	$\mathbb{R} \in [-1,1]$	$\mathbb{R} \in [-1,1]$	$\mathbb{Z} \in [-1,0,1,2,3,4,5,6]$	$\mathbb{R} \in [0,360[$

The values of table 3.4 portray the following information:

- **frameNo**: an integer greater than or equal to zero that describes the position of an image in the sequence of images.
- **t**: a positive real number describing the time of the run in seconds in which the image identified by *frameNo* was taken.
- **speed**: a positive real value that is the current car speed in m/s.
- **steering**: a real value between negative one and one. Positive values represent steering to the right, whereas left values represent steering to the left. The closer the value is to one of the extremes, the stronger the curve performed.
- **accelerating**: a real value between negative one and one. Positive values represent acceleration whereas negative values represent braking.
- **braking**: the opposite of *accelerating*. Positive values represent braking, whereas negative values represent accelerating.
- **currentGear**: an integer between negative one and six, representing the car's gear at the time. The negative gear is the reverse gear.
- **headAngle**: a real number translating the angle at which the driver's head is pointed, clockwise. If the value is zero, the driver is looking perfectly forward.

Lastly, the third file contains the following structure (Table 3.5):

Table 3.5: CSV template for pedestrian data.

pedestrian	frameNo	t	visible	gazeVisible	nearCrosswalk	crossing	distance	groupSize
$\mathbb{N}_0$	$\mathbb{N}_0$	$\mathbb{R}_{>0}$	*true / false*	$\mathbb{R} \in [0,1]$	*true / false*	*true / false*	$\mathbb{R} \in]0,1]$	$\mathbb{N}_0$

The values of table 3.5 convey the following information:

- **pedestrian**: a positive integer that identifies the pedestrian. This number is attributed to pedestrians in the order that they were first seen.
- **frameNo**: an integer greater than or equal to zero that describes the position of an image in the sequence of images.
- **t**: a positive real number describing the time of the run in seconds in which the image identified by *frameNo* was taken.
- **visible**: a boolean value translating whether that pedestrian is visible in the scene.
- **gazeVisible**: a positive real value that describes the direction of the normal vector of this pedestrian's gaze, if present in the scene. If not, it is zero.
- **nearCrosswalk**: a boolean value translating whether that pedestrian is near a crosswalk.
- **crossing**: a boolean value translating whether that pedestrian is crossing the street.
- **distance**: a positive real value that translates the distance at which the pedestrian is from the car. This distance is the nearest possible distance considered from every pixel that displays the pedestrian in the current image.
- **groupSize**: an integer representing the size of the pedestrian's group. If not in a group, this value is zero.

3.5.2 Data Preparation

After compiling data in files, the next step for analysis consists of preparing it for analysis. Separating it in different files provided the advantage of easier visualization of three different domains of study during driving: the visual information that could be obtained, the driver metrics and all pedestrian information. But this meant that all obtained data would be split among such files for each driving run. Thus, it was important to prepare data for analysis by compiling relevant information together.

To create pedestrian history, it was vital to sort data by pedestrian. The obtained file contains separate lines for each pedestrian for each time interval. Thus, data preparation in this front firstly consisted of such sorting by pedestrian number.

Sequentially, driver information was needed for each pedestrian so as to create comparisons between each other's history. Thus, driver information from the second file (Table 3.4) was joined to the sorted pedestrian data so as to create such relations.

To create maps, boolean values were replaced with integers wherein zero means false and one means true.

Prepared data was stored in different files from the ones mentioned in the previous section.

When analyzing results, data was processed using RapidMiner Studio's tools. They allow for quick selection of operators to apply to data imported from the CSV files and also for quick visualization of data in selected graphical formats. For result analysis, several experimental setups were created in RapidMiner Studio (Fig. 3.20), depending on the data task at hand. While some data required only selection of certain lines in a file, other variables needed to be related to other variables in order to get the wanted result format.

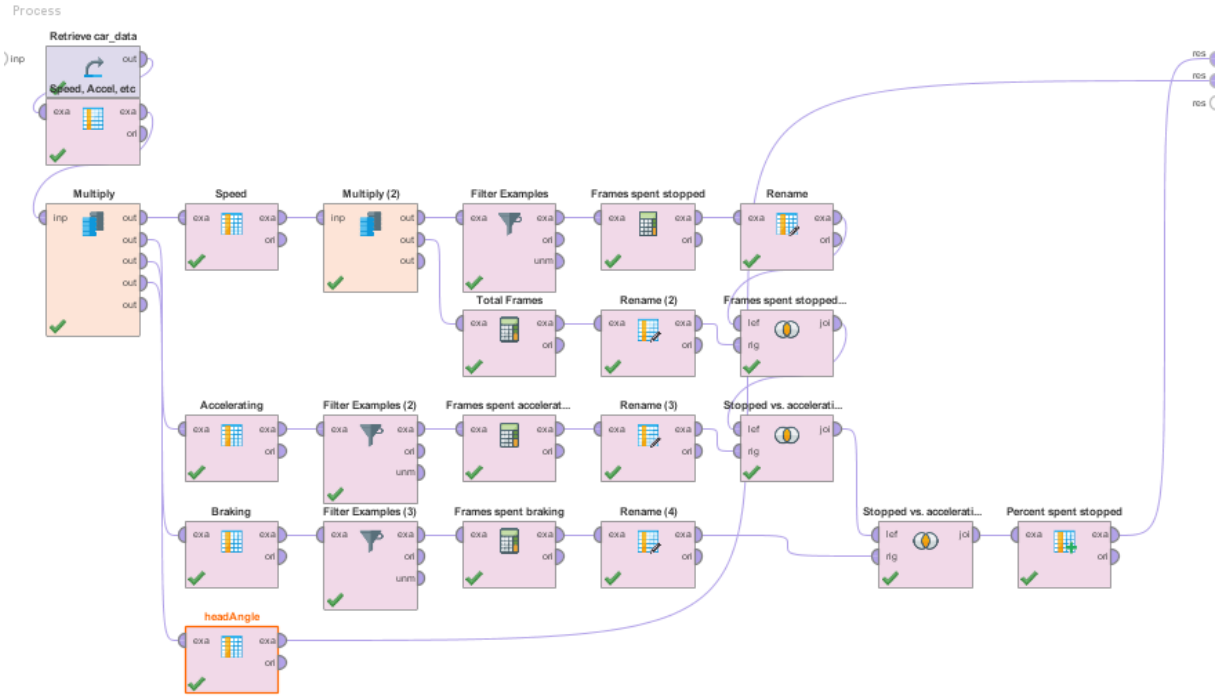


Figure 3.20: The setup needed to obtain speed, acceleration, braking and head angle visualization.

3.6 Research Study

All the previous sections described the needed methodology for gathering and preparing data to analyze results over which conclusions could be inferred. Thus, it was required to conduct several driving experiments with real subjects so as to gather such data from real participants.

The objective of driving experiments would be not only to emulate a real driving experience while in a safe environment but to have the layered methodology alongside such experiments gathering knowledge for conclusions.

The research study that was conducted was intrinsically quantitative. Previous sections described how gathered values were useful for the study at hand. As such, this quantitative approach aims for a statistical analysis at its final phase. This statistical modelling is useful in the sense that it provides objective data over the context, and that conclusions are easily visualized if information is plotted using some kind of visualization tool. It also provides a starting place for a predictive solution there on after.

Naturally, a quantitative analysis of any context follows some assumptions so that the statistical model can be constructed. This leaves conclusions dependent on such assumptions to a certain degree. Nevertheless, this study's aim involved finding common factors that could be pinpointed during driving that influenced driver inference over pedestrians. Quantitative approaches provide the needed methodology for this. Qualitative approaches, on the other hand, would be focused on

the participants' opinions. While it would be very interesting to also obtain data over demographics and direct opinions over human inference in driving scenarios, this approach has been done many times in related work as mentioned in chapter 2.

The experiments consist of having a subject drive a virtual cockpit using VR and a physical racing wheel through a set track. During this track, pedestrians and other elements are encountered. After the experiments data is gathered over their driving and how they act during driver-pedestrian interactions.

Two different experiments are performed per subject. One is a simple track with pedestrians as the only dynamic elements present in the simulation. The other one also features moving cars that compete for space with the subject. While the focus of both experiments is on the pedestrian encounters, the second experiment serves to check if the presence of other dynamic elements affects driver judgment during encounters.

The two tracks are similar between the two experiments, except for the number of elements present. Effectively, subjects go through the same track twice. In the first experiment drivers are still new with the environment and controls. After conducting the first experiment, it is expected that pedestrians reach a certain degree of mastery of them. This means that subjects will learn from the first run. This results in the two experiments not being independent in this regard. Nevertheless, two reactions to the same occurrence will not be exactly the same in any two experiments and drivers. Thus, this crossover effect is somewhat negated. It should, however, be considered in the results.

After a decent sample has been gathered the statistical model can be constructed. This model can be the basis of the answers for the research questions enumerated in chapter 1.

3.6.1 Assumptions

Naturally, due to the environment used for the experiments results will be dependent on certain assumptions, as mentioned previously. The driving experience aims at being the most realistic as possible, but of course it is very different. The controls are somewhat different and some users will suffer from problems due to them and due to the virtual reality setup.

Thus, results have to take into consideration these factors. The main result that should be affected by the setup will be the head movement of drivers. Subjects in VR tend to move enthusiastically and fast, thus making these results have faster changes and more extreme values. One other that should be affected by the setup will be braking values. The brake on the pedal setup is quite dramatic, and brakes very fast and hard. Thus, values will suffer from the same quick variations between any two frames. Nevertheless, these variations will not influence the presence of braking values, just their amplitude.

Moreover, the simulation has certain limitations that have already been mentioned. Pedestrians behave somewhat erratically when approaching crosswalks, especially if surrounded by objects or other pedestrians. Pedestrians are also collidable and have a heavier mass than the cockpit in Unity, which makes collisions very difficult to recover from.

Moving cars are extremely slow and not collidable. This can make subjects frustrated and make braver decisions than what they would do in real settings. Nevertheless, they are warned before the experiments that they should at all times strive to behave as realistic as possible, while still being permitted a margin of error for study.

3.6.2 Expected Results

Preparing a research study allows the anticipation of some answers to the research questions. Some results can be somewhat expected, by intuition or common occurrence. Nevertheless, not only obvious results should be considered or anticipated.

In terms of the research study, some results can be expected from drivers. If experiments are well conducted and feel somewhat realistic, drivers will react as though they are in the real world. For this reason it is valid to assume that:

- Drivers will follow road rules.
- Drivers will stop to let pedestrians cross.
- Drivers will stop at intersections to check if there are any incoming cars or pedestrians.
- Drivers will aim not to hit any obstacles, be them static or dynamic.

These presumptions are somewhat obvious. Drivers will not easily be eager to be in danger or to put pedestrians in danger. They should strive to make their experience as smooth as possible, both in the real world as well as virtually. These presumptions will be checked through results. It is expected that the latter corroborate the former, and with little uncertainty.

However, these do not provide much input to the study of driver-pedestrian interactions. Simply confirming them will not provide enough to conclude over interesting perceptions that affect driving in pedestrian encounters.

Therefore, it is vital to confirm some other theories. Expected results about driver-pedestrian interactions include:

- Drivers will be more willing to stop for pedestrians that are closer to the car when crossing. They should be paying more attention to obstacles close to the car, and strive not to hit any of them **(ER1)**.
- Drivers will be more willing to stop for pedestrians that they have seen for less time. If a pedestrian suddenly appears within the field of view of the driver and is crossing the road, this should trigger a faster reaction to stop than a pedestrian that is simply stood by the sidewalk waiting to cross **(ER2)**.
- Drivers will stop more easily for pedestrians that are accompanied by other ones if waiting to cross the road. This effect should be even higher should there be a group on each side of the road **(ER3)**.
- Drivers will pay closer attention to their surroundings and drive more carefully if there are other cars competing for space on the road. The crowded environment could be blamed for this, and should make driving slower and stopping more frequent **(ER4)**.

- Drivers will stop at stop signs even if there are no people incoming. They should stop more frequently in intersections if there are pedestrians around **(ER5)**.
- Drivers will yield to people that made eye contact or portrayed other significant body language to them while waiting to cross the road **(ER6)**.
- Drivers will not stop as often if pedestrians that are walking towards the crosswalk take a long time or portray a calm stance **(ER7)**.

The results and statistical data obtained from the experiments should be the only factors to take into consideration when confirming these affirmations. Given this, the results that should be looked at in order to confirm each expectation should be:

- **(ER1)**: braking, speed and pedestrian distance in pedestrians in a *crossing* state.
- **(ER2)**: the time interval in which each pedestrian is *visible*, in relation to braking and speed.
- **(ER3)**: the group size for pedestrians whose state is *nearCrosswalk*, in relation to braking and speed.
- **(ER4)**: speed, accelerating, head movement and braking in images where there is another moving car present, and distance to such car.
- **(ER5)**: sections in which a stop sign is presence, and possible correlation with speed and braking. Also, images where both stop signs and pedestrians are present and a similar correlation, if it exists.
- **(ER6)**: intervals where pedestrians are in a *nearCrosswalk* state for a long time and correlation with speed and braking. A comparison of such long intervals to ones where they are short should be made.
- **(ER7)**: drivers' head angle and the normal direction of each pedestrian gazing at the driver, and relation with speed and braking.

3.7 Predicting Pedestrian Intentions

The methodology for gathering data can give us a look into the variables needed to make a judgment over factors that influence driver-pedestrian interactions on a visual and contextual level. Although this study's results focus mainly on the statistical model created through the exploration of such data, it is important to mention that such data can make way for predicting outcomes of scenarios.

It is important in simulating driver-pedestrian interactions a tool for predicting pedestrian behavior is in place in real time. With this, one could achieve a reasonable set of behaviors that are, through the proper methodology and training, quite similar to real drivers' predictions. Data gathered through experiments (a rather large sample) can provide the needed information for training and testing such a predictive model.

Several different predictive models have been mentioned in chapter 2. Decision Trees could provide a simple predictive tool that is rule-based and can provide direct visualization over which variables play the biggest role in inferring pedestrian intentions. SVMs have the ability of turning

this rather complex and non-linearly separable problem into one that is separable, due to the nature of the algorithm itself. It is hard to say which one would provide the best results. One or the other may provide more reliable outcomes with proper training. Besides these two, deep networks can be used to predict intentions as well. On the other hand, they require a large dataset in order to achieve good predictive results, which, with the methodology at hand, would prove quite arduous since numerous experiments would have to be performed. Nevertheless, they are an interesting approach to the predictive problem, and can be coupled with other deep network methodologies mentioned previously, such as the pose detection problem and the segmentation problem.

In order to construct such a predictive tool, the dataset would require splitting into a training dataset and a testing one. The training dataset would teach the model about the relevant variables in play and associated outcomes. The testing dataset would be used to validate such results, and to fine-tune the predictions. Predictions are associated with labels. In this case, a label would be the confidence value that the pedestrian is going to cross. However, these may or may not correspond to the real labels that are expected. In order to validate such prediction results confusion matrices and ROC charts are the tool to use.

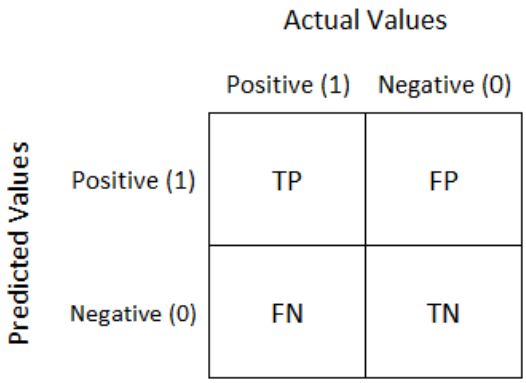


Figure 3.21: The confusion matrix.

According to the matrix (Fig 3.21), values that are predicted may be positive or negative. This is similar with real values. When predicted values are in correspondence with real ones, they are called true positives and negatives (TP, TN). Conversely, when they do not correspond they are defined as false positives and negatives (FP, FN). The proportions of each of these four corners of the matrix will provide knowledge about the veracity of output results and the predictive capabilities of the models.

Two of the most important metrics to be used with these values are the sensitivity (or True Positive Rate) and specificity (or True Negative Rate). They are defined as following:

$$Sensitivity: \quad TPR = \frac{TP}{TP+FN} \tag{3.2}$$

$$Specificity: \quad TNR = \frac{TN}{TN+FP} \tag{3.3}$$

These values can be calculated and plotted as a ROC chart (in fact, one uses the inverse of the specificity as the x axis). This chart (Fig. 3.22) gives us information about the predictive capabilities of the model. One should strive to have the resulting curve to be as close to the top left corner as possible.

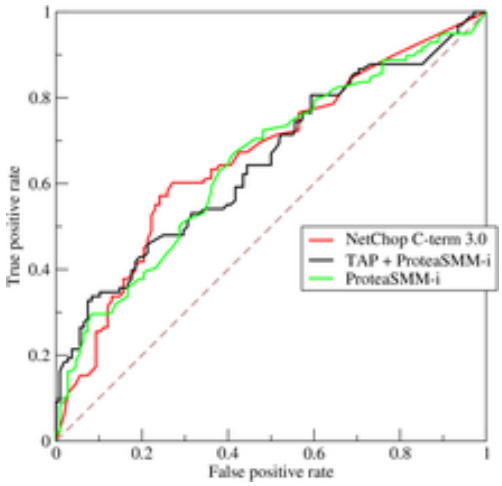


Figure 3.22: An example ROC chart.

These two tools allows for the fine-tuning of the prediction models as well as giving insight into their reliability. They should be used for all prediction tasks in this scope.

3.7.1 Summary

The methodology that was created focused on the process of extracting information from scenarios and processing it into a dataset to serve as a basis for prediction.

The methodology was implemented in a pre-existing Unity3D environment and an external simulation, whose modularity facilitated development. The environment represented a three-dimensional city to serve as the basis for the testing of the pipeline.

Visual and driving data is collected from scenarios and is then processed simultaneously into a dataset consisting of visual and contextual information in those scenarios.

Visual data is processed using different processes of image segmentation, as well as the creation of depth and normal maps. All these allowed for the calculation of other relevant metrics, such as pedestrian gaze presence and pedestrian groups. Contextual information is saved throughout the driving runs and stored for each pedestrian.

Data was stored and compiled in different CSV files that are the basis for a predictive model, to be validated using the regular metrics.

Chapter 4

Experiment, Results and Discussion

Appropriate context into the necessity of performing experiments with subjects has been given in the previous section. This section aims at detailing the process of experiments, as well its results and a discussion of their meaning.

4.1 Experiments

The Unity environment that was used for development was useful for its modular characteristics, as well as the ease of implementation of image segmentation techniques and other algorithms for processing image and contextual data. To obtain results that could come into agreement with expected results it was necessary that experiments were performed over a controlled environment.

Thus, due to the openness of the existing virtual city landscape, it was necessary to restrict experiments to certain areas. This was because there needed to be a controlled study of pedestrian encounters. If all encounters were random and different for every subject, then all contextual information about them would also be different. This would render extracting any significant conclusions quite difficult, since every result would have to take into consideration where and how it took place.

Thus, a fixed itinerary was created for the experiments. All experiments follow this itinerary. Throughout the journey, a driver has the liberty to decide how to react to pedestrian encounters, but these encounters are fixed in place (they always happen in the same places for every experiment) and in number of pedestrians and their starting position (so as to assure the visibility and group investigations).

Experiments were divided into two parts. In a first experiment, the subject drives through the itinerary only encountering static obstacles like clutter on the sidewalk, stop signs and parked cars, but also pedestrians that cross the street in those fixed spots. No cars are present in this first part. In the second experiment, cars are also present in the environment and follow fixed routes that compete with the subject's itinerary.

4.1.1 Preparation

Given the openness of the Unity environment, only a small area was permitted to be part of the route for experiments. The cockpit is inserted at the beginning of the track (Fig. 4.1), where subjects get accustomed to VR and pedal controls and the insides of the cockpit.

Figure 4.1: The cockpit as seen from the outside.

The existing scene for pedestrian generation and handling consisted of random spots around town that could spawn pedestrians, as well as arbitrary spots that pedestrians would see as goals and strive to reach. When pedestrians reached a goal, they would find another one and set it as their new goal. This created an infinite stream of pedestrians walking around town competing for space. Thus, it proved to be a dynamic environment (although not a controlled one). In a first attempt to run experiments, this pedestrian scene was used. It distributed fifty pedestrians equally into those sources and attributed to them those randomly placed goals. This proved confusing for subjects, since most times they would have a pedestrian in sight or close to the car and would thus be stopped or cautious most of the time. Moreover, the presence of fifty pedestrians in the simulation made performance insufferable, and subjects had much more time to think about their actions than they would otherwise. This situation demanded that results from this first batch of experiments would be discarded since they didn't prove to be realistic enough for research.

This led to the remaking of the first track. Instead of having a constant presence of pedestrians freely walking around town, their spawning was limited. When the driver's car would reach a certain fixed point within the track, pedestrians would spawn in nearby intersections. This should only work for the cockpit car, so that incoming simulation cars do not affect pedestrian presence. Trigger walls (Fig. 4.2) were created to be configurable in the number of pedestrians spawned and the goals associated with them. Thus, by having the same settings in all experiments it was guaranteed that pedestrians would follow similar behaviors. Behaviors were not always the same because of different subjects' driving, but also the perception system in play. Despawn triggers were also created, but their goal was to remove created pedestrians from the scene when they would not be useful anymore, so that performance would not be hindered throughout the experiment.

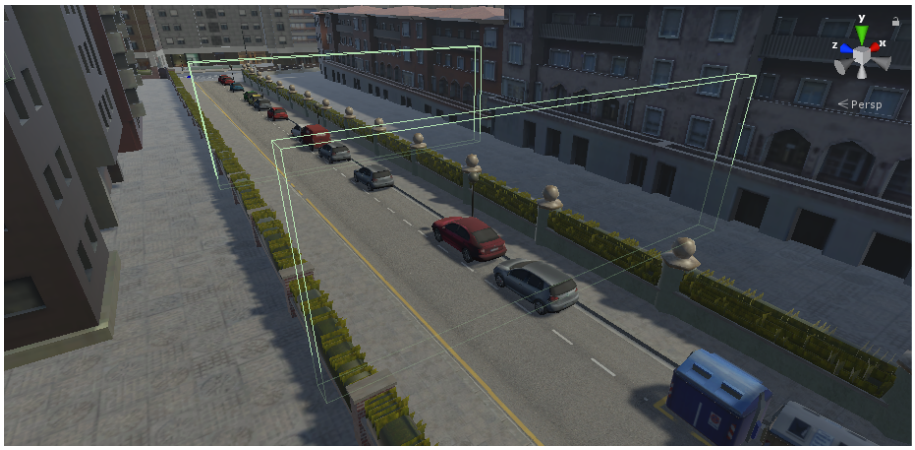


Figure 4.2: Pedestrian spawn triggers laid out on the track.

In order to make experiments controlled, the track at hand also needed to be controlled in the movement possibilities that are available to subjects. It was undesirable that in such a large environment a subject could simply drive off into parts that did not feature triggers, making pedestrian encounters non-existent. To limit this, collidable walls (Fig. 4.3) were installed on the sides of the track to make subjects obey the track direction at all times. These walls feature a large prohibition symbol, which is easily perceptible to subjects as an area that is off-limits. Of course, this allows presents itself as something that may influence drivers' behaviors and the overall study. In the real world, these objects do not exist and all objects are an integral part of the environment they are in. These walls stand out in the scene as foreign objects, and may affect driver judgment. Nevertheless, these non-diegetic objects facilitate the experiments without affecting pedestrian encounters significantly.

Figure 4.3: Guiding prohibition signs that signal the user not to drive this way.

The track was designed to feature several straight sections and few turns and intersections where users would encounter pedestrians. As mentioned, these encounters happened in selected areas. Thus, the track features the following sections (Fig. 4.4):

- **Start**: The starting point of the experiment. The cockpit car is spawned in this area and this is where users get familiar with controls and begin driving.
- **Intersection 1**: After driving ahead from the starting position, users are presented with a four-way intersection where two roads are blocked by prohibition signs. In this intersection three pedestrians cross the street at the same time.
- **Intersection 2**: Driving to the end of the road leads users to a three-way intersection where two pedestrians cross the street at different times. Users are forced to turn left, into a two-way street. In the second experiment, this road is also filled with incoming vehicles.
- **Intersection 3**: Users are forced to turn left into an intersection where some pedestrians are waiting indefinitely to cross.
- **Intersection 4**: Driving to the end of the road leads the driver into a crowded four way intersection where the only possible exit is to turn right.
- **Intersection 5**: Users face an intersection where a stop sign is present. No pedestrians cross this street. In the second experiment, this road is crowded with cars competing for space with the driver.
- **Section 6**: After driving half the road to the finish line, a pedestrian pops into scene and tests drivers' reactions in an encounter outside of a crosswalk.
- **Finish**: The finish line is signaled by a white cube floating in mid-air.

Each section aims at testing different types of pedestrian encounters and associated driver actions. In every intersection there is a different number of pedestrians that spawn and different behaviors that they perform.

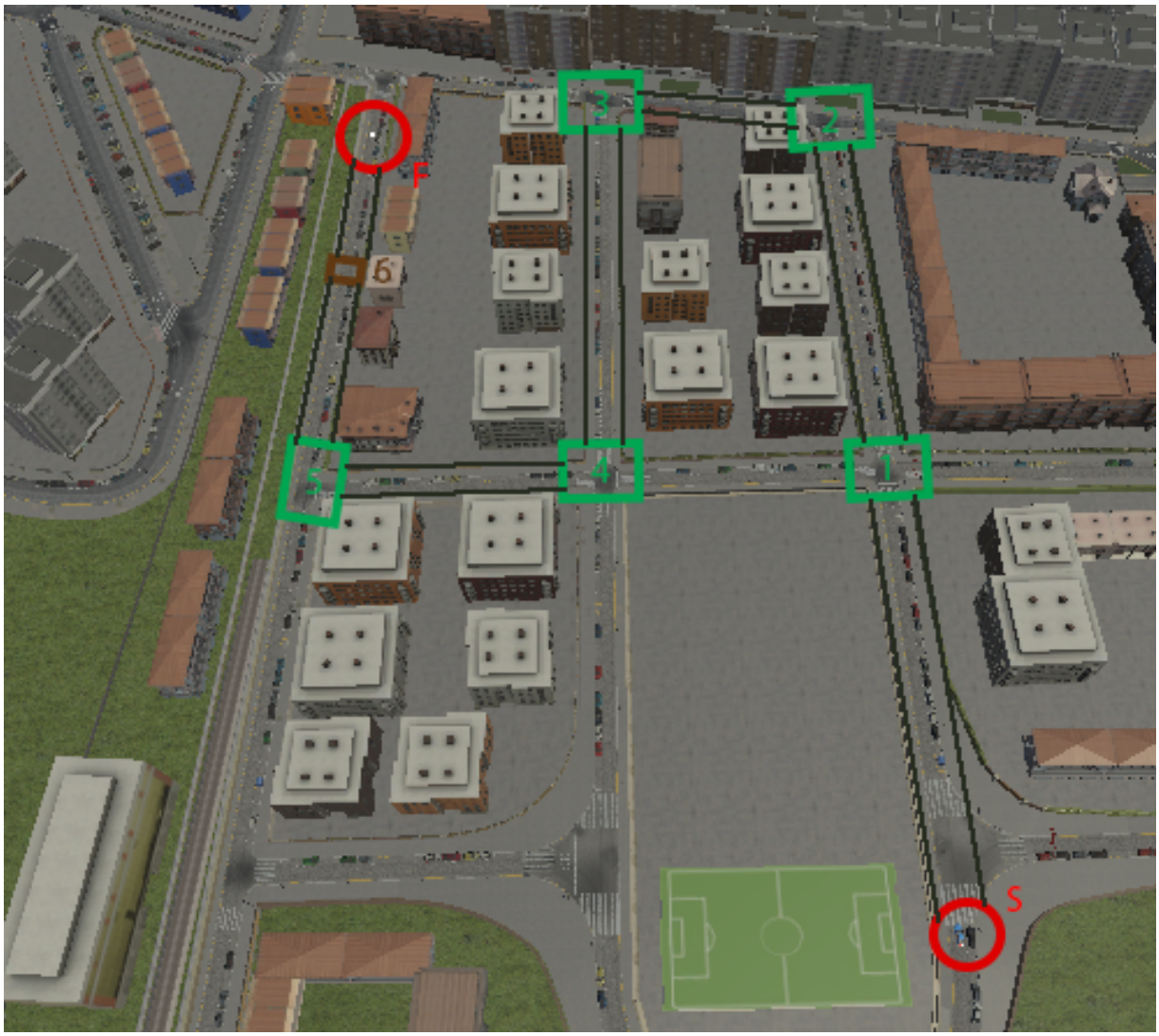


Figure 4.4: The map of the itinerary.

Intersection 1 (Fig. 4.5) aims at familiarizing drivers with the concept of pedestrian encounters. When approaching first crosswalk, three pedestrians cross the street at the same time. This encounter is quick and pedestrians are usually already crossing when the driver reaches the crosswalk. Pedestrians go one by one, and are not grouped. Thus, this section tests drivers' actions to individual pedestrian crossings.

Figure 4.5: A crowded first intersection.

Intersection 2 (Fig. 4.6) has as its goal testing drivers' reactions to wide crosswalks in which pedestrians take a long time to reach the crosswalk. Drivers that disregard their surroundings will not take notice of incoming pedestrians. Those who driver more carefully will see that pedestrians are trying to cross. While pedestrians coming from the right are visible for a long time before reaching the crosswalk, the pedestrian coming from the left is not visible until it is very close to the crosswalk. Drivers should also look to their left and right before going into the next section.

Figure 4.6: The second intersection of the itinerary.

Intersection 3 (Fig. 4.7) tests the effect of pedestrian visual contact with drivers. Pedestrians are standing on the crosswalk waiting to cross. They look directly at the car. Drivers that are careful will notice pedestrians' intentions to cross and should stop prior to reaching the intersection. They should also look to establish visual contact with the pedestrians to read their intentions.

Figure 4.7: Intersection number three and pedestrians waiting to cross.

Intersection 4 (Fig. 4.8) is reached after a long straight segment of the track. In that segment, drivers typically accelerate to high speeds. When they reach the crosswalk they are presented with various groups of pedestrians crossing in two different crosswalks in the intersection. This crowded environment aims at testing the effect of pedestrian groups in driver judgment. Pedestrian groups cross the road at the same time.

Figure 4.8: Different groups of pedestrians wanting to cross at the fourth intersection.

Intersection 5 (Fig. 4.9) presents the subject with an intersection where there are no pedestrians to be seen. Instead, a stop sign forces drivers to take caution into the next right turn. Drivers that are less careful will ignore this sign.

Figure 4.9: A stop sign before an intersection.

Section 6 (Fig. 4.10) is present in the final stretch of the track. When driving down this stretch, drivers typically accelerate if no obstacle is in sight. However, when reaching this section a pedestrian that is walking the sidewalk suddenly decides to cross the street, even without a crosswalk. This section aims at evaluating drivers' responses to pedestrians that are suddenly crossing.

Figure 4.10: A pedestrian that unexpectedly crosses the street at the final stretch of the itinerary.

The second experiment featured cars in key points of the track. After the second intersection, cars are incoming (Fig. 4.11) from the left and right and distract drivers about incoming pedestrians. This was done in order to evaluate how the presence of more dynamic objects in the scene could affect driver judgment.

Figure 4.11: Incoming cars on the second intersection.

4.1.2 Experimental Protocol

Conducting experiments meant the necessity of establishing a protocol so that every experiment would be conducted smoothly and similarly. People were invited to participate in the experiment if they so wished. During the process, such protocol took place. The protocol was established as such:

- **Preparation**: Before conducting experiments, all the material was gathered in the work station in order to start the experiments with all that was required. This material was the VR headset, a computer capable of running the environment, and the racing wheel and pedal set.
- **Introducing the subject**: Subjects were greeted and introduced to the station where all the material was gathered. It was then explained to them the usage of each part of the material. The VR headset and base stations and their way of operating was also described in detail.
- **Goal explanation**: Subjects were introduced to the goal of the study and how the material presented to them would be used in the experiments. While expected results were not described to the subjects, they were informed of what information about their experiment was being gathered and assured that no personal information or feedback after the experiment would be needed, since it was a quantitative study.
- **Consent form**: If subjects showed interest in partaking in the experiments and understood what was being studied, they were presented a consent form. Signing this consent form formalized the subject's interest and willingness to participate in the experiment. The form (in annex A) did not imply that subjects were bound to finish the experiment and they were informed that they could quit any time during the experiment if they so wished.
- **Control explanation**: Driving and headset controls were explained in detail to subjects. They were first shown the racing wheel and pedals, and how they differ from the controls of a real car. The headset was also installed on the subject's head. Subjects were asked for

their level of comfort in virtual reality, and it was reiterated that they could leave at any time if controls felt uncomfortable.

- **Learning the controls**: Subjects were put in a test track in which they could learn how to control the car and their vision inside the virtual environment. They were told to stop when they felt accustomed to the controls.
- **Running the first experiment**: Subjects were put inside the main track and were guided throughout it. Although signs along the track guided them through it, they were also instructed over the turns that they should make. Subjects were never instructed to brake or change their driving, but only guided through each turn.
- **Running the second experiment**: Subjects were told to repeat the experiment, but were informed that the new one would feature other cars.
- **Finishing the experiment**: The headset was removed from the subject's head and it was made sure that all necessary data had been collected.
- **Thanking the subject**: After finishing the experiments, subjects were thanked and it was reiterated that no personal information would be used, only the results from the environment.

This protocol made sure that all experiments ran smoothly. No participant decided to leave and no participant felt uncomfortable during the experiments. In total, twenty people participated in the study. One participant was asked if they felt comfortable with having their picture taken in order to document the experiment setup, to which he agreed (Fig. 4.12).

Figure 4.12: A subject undergoing the experiment.

4.2 Results

After running the experiments, large amounts of image and text data was created to be used for analysis. These two types of information are the basis for the results and the conclusions to be extracted from them.

The first step upon beginning data analysis was to explore the data that was gathered. This meant extracting some general knowledge about the experiments as a whole. This was done to get all needed context in order to begin diving in deeper in the results.

4.2.1 Data Exploration

Concerning the twenty experiments conducted, a total of 15054 images were captured throughout them. Every single image was coupled with a visibility table, and pedestrian information.

Although the track in the experiments was the same for all subjects, their driving experience and time to complete the track was different for every one. The shortest run lasted only 2.58 minutes, while the longest lasted for over six minutes. On average, an experiment took roughly four minutes to complete. Experiments in which there were other cars in the scene lasted for over a minute longer, on average, than their counterparts without them. These variations in experiment times do not influence the results in a significant way. Each subject took their time to get accustomed to the controls and to the scene in general. This lead to such differences in experiment duration. Besides this, runs in which there were cars meant that subjects had to respect other cars' driving. The cars in the scene were very slow, which slowed down drivers.

Of all captured images only some depicted pedestrian encounters. This smaller subset was what was used for extraction of conclusions about driver-pedestrian interaction. In total, there were recorded about 800 pedestrian encounters, which meant roughly twenty pedestrian encounters per experiment. As mentioned previously, not all encounters happened under the same circumstances. Some of the pedestrians were grouped up, some were slow and some others looked at the car. Only 45% of all captured frames depicted pedestrian encounters.

Regarding driving data, it varied between experiments. While some experiments had the subjects drive carelessly and at high speeds, some contained the opposite cases. Figure 4.13 shows information about drivers' inclinations toward accelerating and braking. Drivers mostly accelerated throughout the experiments, and were much more reluctant to braking in comparison. This might be because many drivers chose to simply let the car speed down instead of hard braking at crosswalks. Nevertheless, every experiment contained some time intervals in which the subject was stopped at a crosswalk in order to let pedestrians cross.

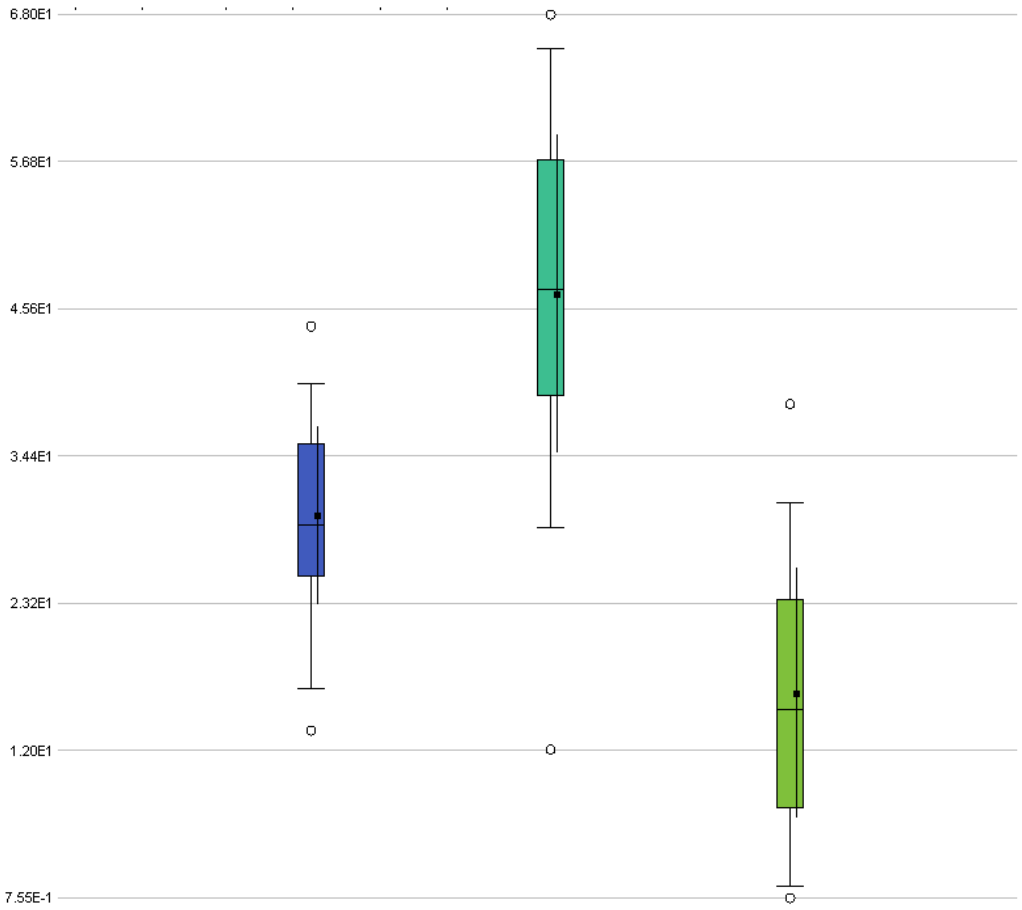


Figure 4.13: Percentiles of frames in which the subject was stopped (dark blue), accelerating (cyan) or braking (green). Values on the left are in percentage in scientific notation.

In order to analyze where a driver's gaze was headed, it was important to gather data about their head movement. The way it was stored recorded only their horizontal angle, and ignored up and down motion. An angle of zero meant the driver was looking perfectly forward, while values over zero correspond to the intensity of their head rotation, clockwise. Thus, values over 345º and values under 15º generally meant the driver was looking ahead, while values close to 270º or 90º generally meant the driver was looking at a certain side. Figure 4.14 shows insight into every recorded angle throughout the experiments. Drivers mostly looked forward, but it is possible to visualize when drivers looked at a side or behind them. These values do not, however, transmit the entire information about the subjects' gaze. As the HMD's field of vision is quite small, subjects usually prefer to turn their head as they want to focus on elements. However, in a perfect scenario this gaze should have also accounted for the eye movement in order to be quite exact. Nevertheless, these values provide a rough estimate of where the driver was looking at.

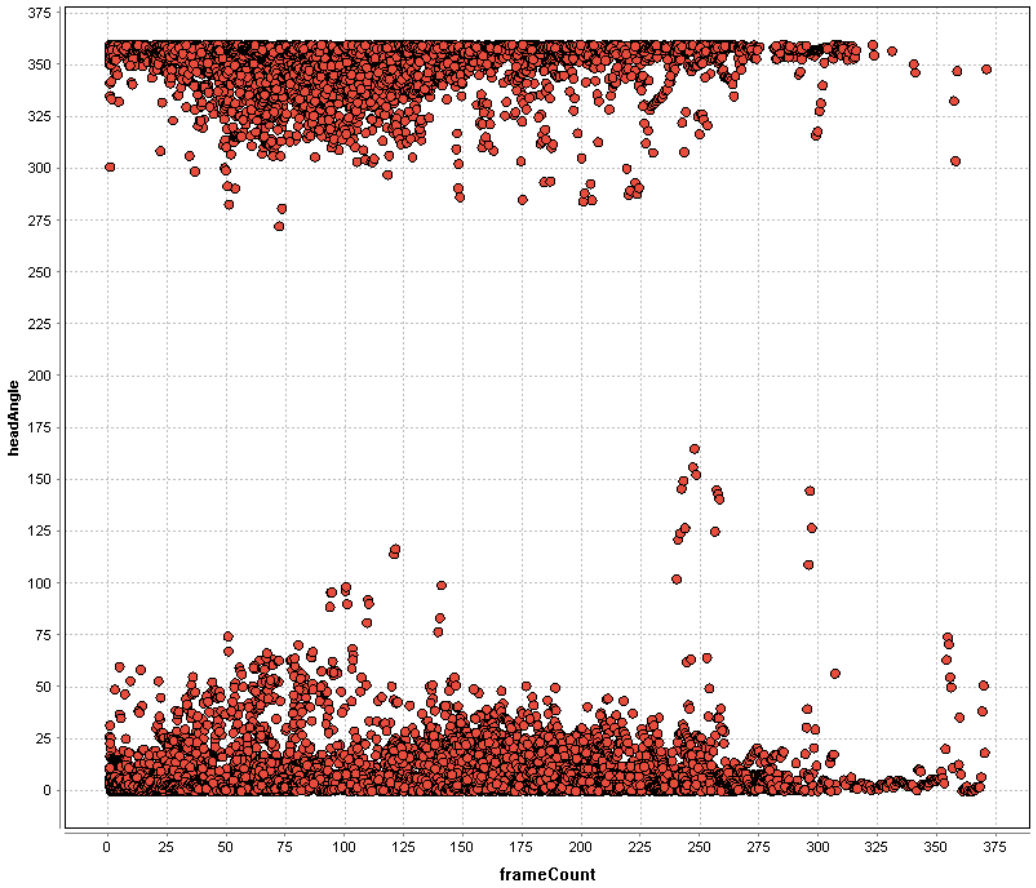


Figure 4.14: Scatter plot of head direction values, for every frame captured.

Regarding pedestrian encounters, it was important to record for how long pedestrians were in sight in relation to the driver. Some encounters were very fast paced while others were slow. This meant a difference in times in which there was visibility of a certain pedestrian (Fig. 4.15). Thus, information was extracted from the pedestrian maps to be used for this analysis. Besides visibility, a pedestrian's state at any time was also extracted. Frames in which a pedestrian was crossing the road were almost as frequent as frames in which the pedestrian is simply walking down the road.

Only some encounters aimed at finding out how visibility time changes drivers' actions. Thus, almost all pedestrians were visible for roughly the same time (around five to fifteen seconds). Some were on screen for much longer. This also changed for every experiment, since some subjects were much more lenient on letting the pedestrian cross than others.

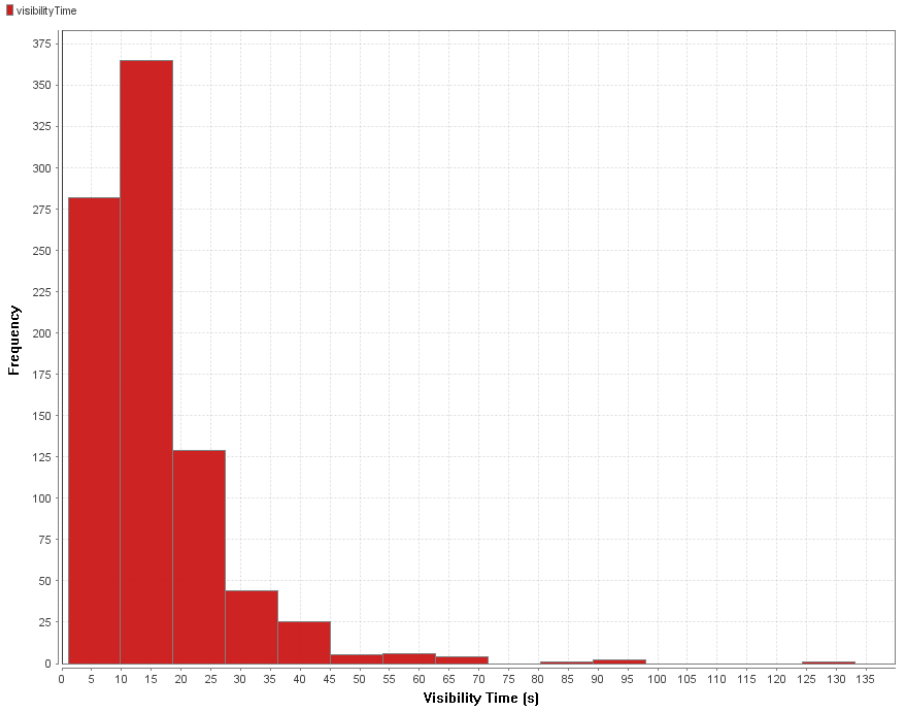


Figure 4.15: Histogram of how long a pedestrian remained in sight for all experiments.

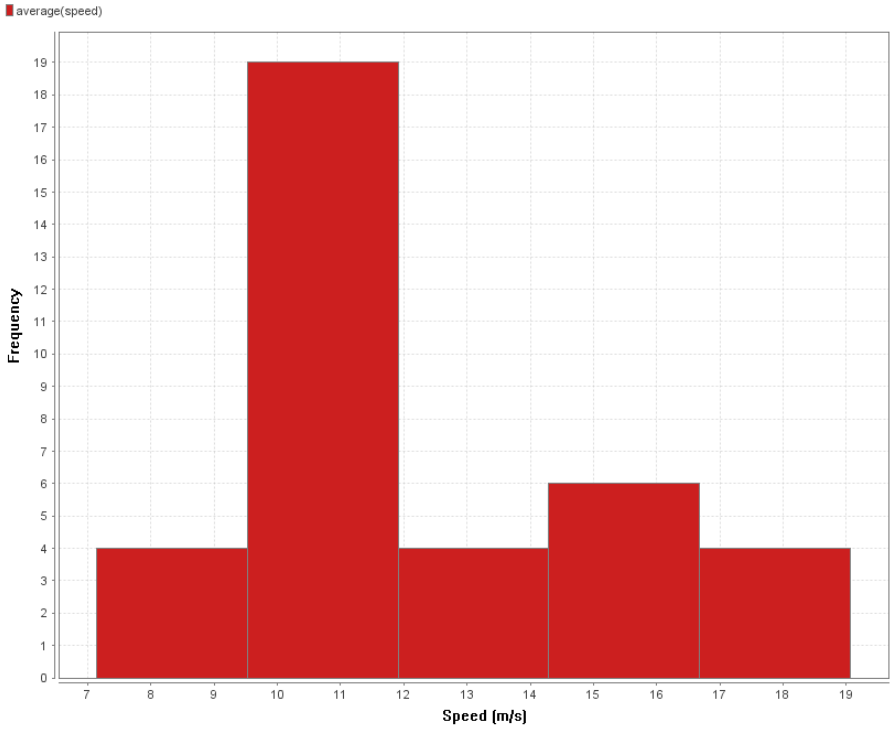


Figure 4.16: Histogram of the average speed in the experiment, for all experiments.

Depending on the subject who was performing the experiment, overall speed changed frequently. Figure 4.16 shows that most drivers maintained a similar level of speed throughout the

experiments. However, a portion of subjects also drove quite fast. This speed also translated into the encounters. In sections where there were no pedestrians in sight drivers drove at much higher speeds than otherwise (Fig. 4.17). This analysis is present in annex B. Driver's drove on average 5.77 m/s faster in when there were no pedestrians in sight. In such sections, drivers' speeds also varied a lot more, with a standard deviation of 4.76 versus one of 1.96.

In sections where there were pedestrians in sight, average speed was much lower for those where pedestrians were crossing the road compared to those where pedestrians were on the sidewalk.

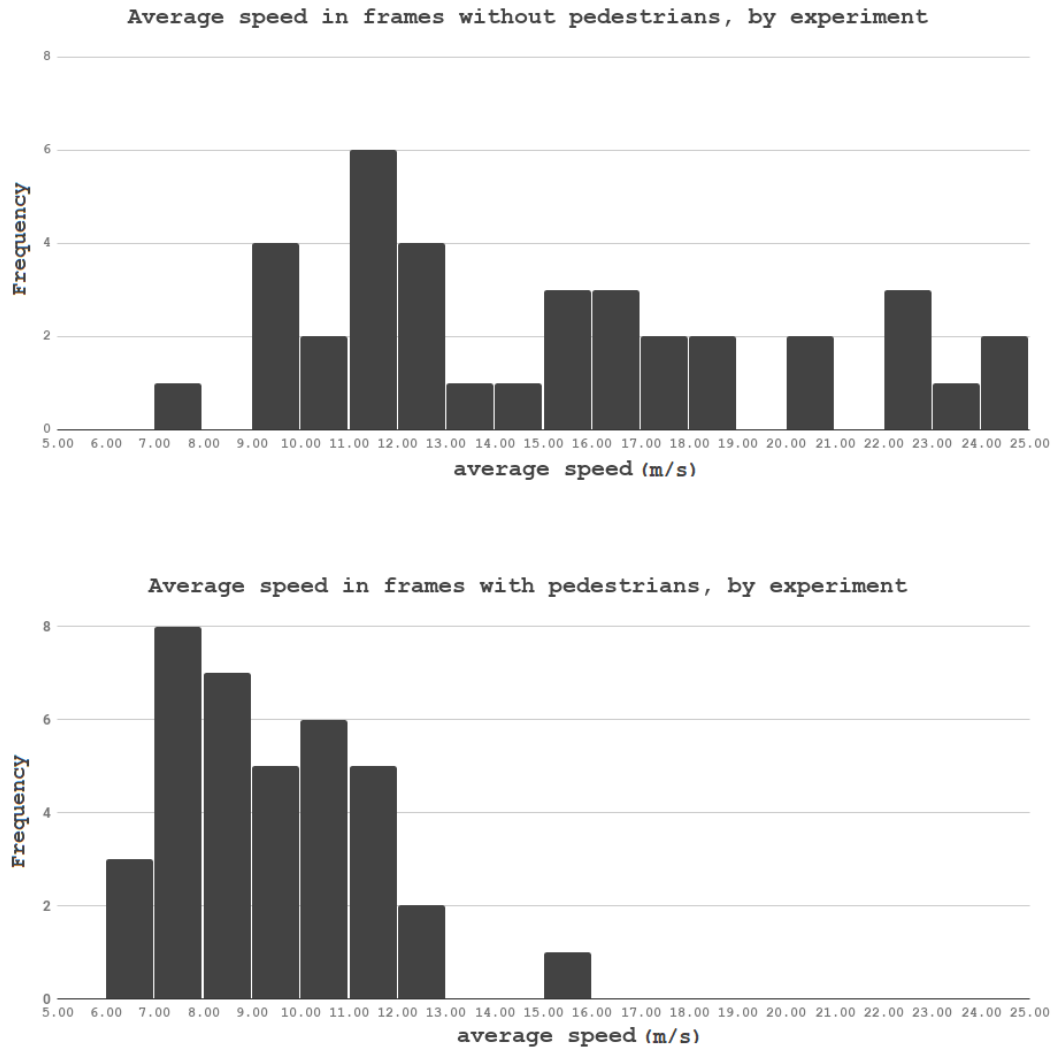
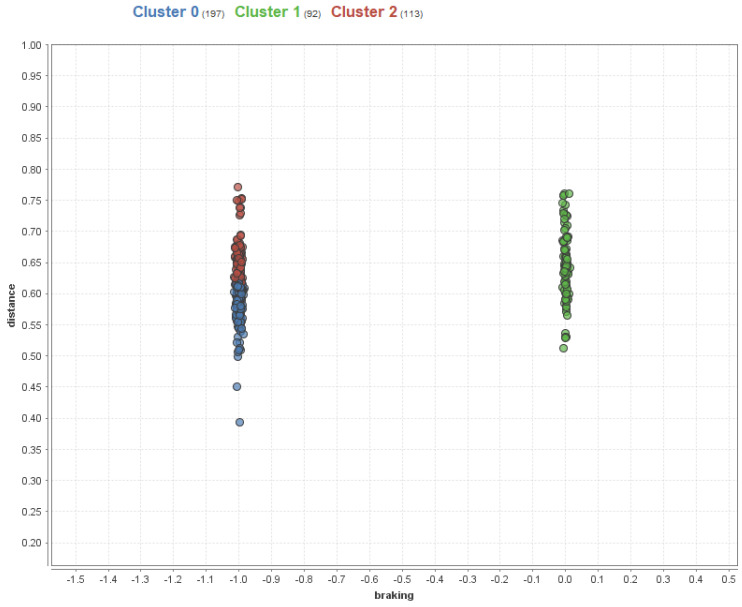


Figure 4.17: Average speed by experiment, in all frames without pedestrians in sight (above) and for all frames with pedestrians in sight (below).

ER1: *Drivers will be more willing to stop for pedestrians that are closer to the car when crossing.*

An experimental setup was created using RapidMiner Studio in order to find out how each driver preferred to react to different distances to a crossing pedestrian. Drivers' actions provide insight into how those distances influenced their braking and speed at crosswalk scenarios. For this, the set of frames in which there were pedestrians crossing the road was used. Then, drivers' braking and the pedestrians' distance to the car was compared. Results can be visualized in figure 4.18. Since drivers' braking actions were quite distinctly separated into different groups, a clustering algorithm was applied to the subset. K-Medoid was chosen because it handles outliers and differences in data density much better than its K-Means counterpart. Three clusters were used. The centroids of these clusters provide input into three different driver profiles:

- Some drivers did not brake at all when reaching the crosswalk. They usually kept a bigger distance to the crosswalk than drivers who did brake (cluster 1).
- Some drivers decided to brake when they were very close to the pedestrian. This was the majority of cases (cluster 0).
- Some drivers chose to brake but at a much bigger distance to the crosswalk (cluster 2). This was seen often during the experiments.



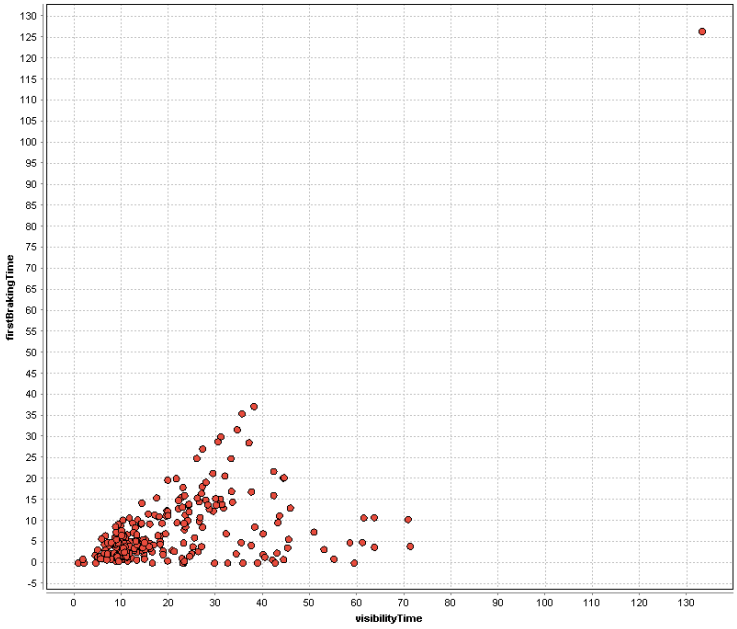
Cluster	distance	braking
Cluster 0	0.580	-1
Cluster 1	0.655	0
Cluster 2	0.651	-1

Figure 4.18: Driver preferences for yielding to pedestrians according to their distance, in clusters.

ER2: *Drivers will be more willing to stop for pedestrians that they have seen for less time.*

Pedestrian maps were used as input to a RapidMiner Studio, as well as driving data for every frame in which those pedestrians were visible. This setup provides insight into the relation between seeing a pedestrian for long before having them cross the road and braking to let them cross. The time for the driver to brake after seeing the pedestrian for the first time was calculated, as well as the total time in which that pedestrian was visible. A correlation between the two might mean that drivers feel much more inclined not to yield to pedestrians that they have seen for a long time.

Figure 4.19 visualizes the scatter plot obtained from those attributes. As can be seen by the shape of the plot, the points are in positions that can indicate the presence of a correlation between the two. The correlation matrix extracted from the plot shows that there is a correlation index of 0.545 (p-value < .00001). Although this value does not mean there is a strong correlation between visibility time and time to brake, it indicates that these factors are in some way related. If such correlation exists, it is a positive one. For many cases, drivers chose to brake much faster and harder if they had not seen the pedestrian for long. This is similar to what was expected as a result.



Attributes	visibilityTime	firstBrakingTime
visibilityTime	1	0.545
firstBrakingTime	0.545	1

Figure 4.19: Scatter plot and correlation table between total visibility time and time to break after seeing a pedestrian for the first time.

ER3: *Drivers will stop more easily for pedestrians that are accompanied by other ones if waiting to cross the road.*

The same setup for the previous result was used in this scenario. However, instead of relating visibility of pedestrians with time to brake, it was swapped out with the number of pedestrians that were grouped up. Most of the pedestrians were alone in the encounters throughout the experiments. However, some encounters featured the presence of some groups.

Results were laid out in a scatter plot that relates time to brake to the size of a group (Fig. 4.20, 4.21). Drivers' actions were much more similar when the group size was bigger. However, the sample size for bigger groups was also smaller. Averaging out the times to braking for each group lead to the second plot in figure. It shows that on average drivers took much less time to brake when group size was bigger, showing a possible negative correlation between the two factors.

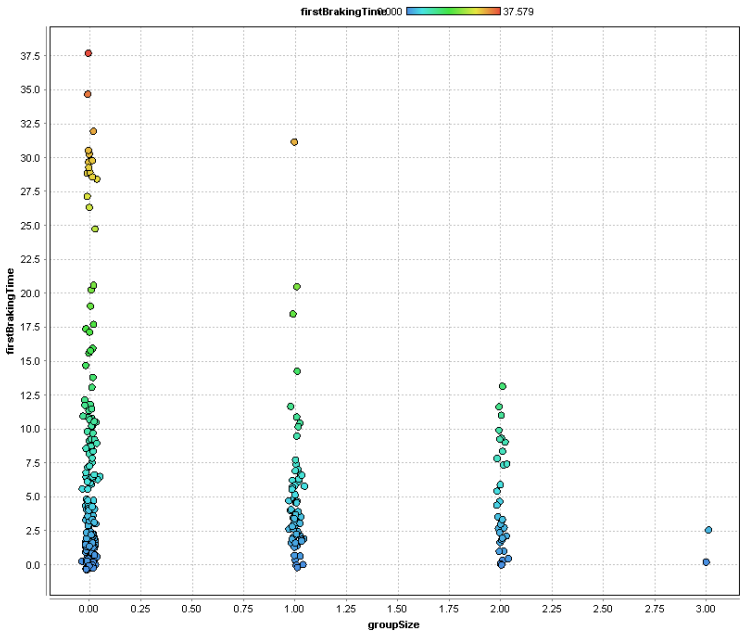
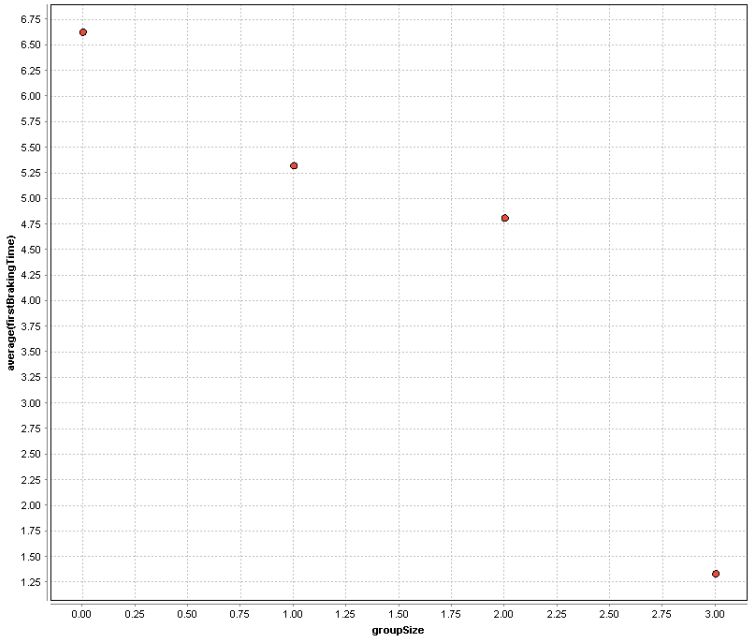


Figure 4.20: Influence of group size on brake times (scatter plot).



Row No.	groupSize	average(firstBrakingTime)
1	0	6.631
2	1	5.326
3	2	4.814
4	3	1.338

Figure 4.21: Influence of group size on brake times (averages).

ER4: *Drivers will pay closer attention to their surroundings and drive more carefully if there are other cars competing for space on the road.*

Driver speed and braking actions were used in order to investigate how other cars affect drivers. As subjects underwent two similar experiments in which only one contained cars, their driving was much different in those that did.

The subset of frames in which there were other cars in sight was separated from those where there were not. Speed values for every one of those frames was gathered, and a histogram of such values was created to better visualize their distribution.

Results on figure 4.22 show that drivers chose to drive at much lower speeds when there were other cars in sight. Although drivers' speeds were quite similar on the lower end of the spectrum, they were less frequent at higher values if there were cars in sight.

Driver head motion was also much more varied when there were other cars in the road (Fig. 4.23). Drivers looked to the sides much more frequently in these cases. In both cases the head angle values were normalized. When cars were not present, this attribute had a standard deviation of 1.014, and when they were this value was of 1.315. This means that in general drivers' head motion was more frequent when cars were in sight.

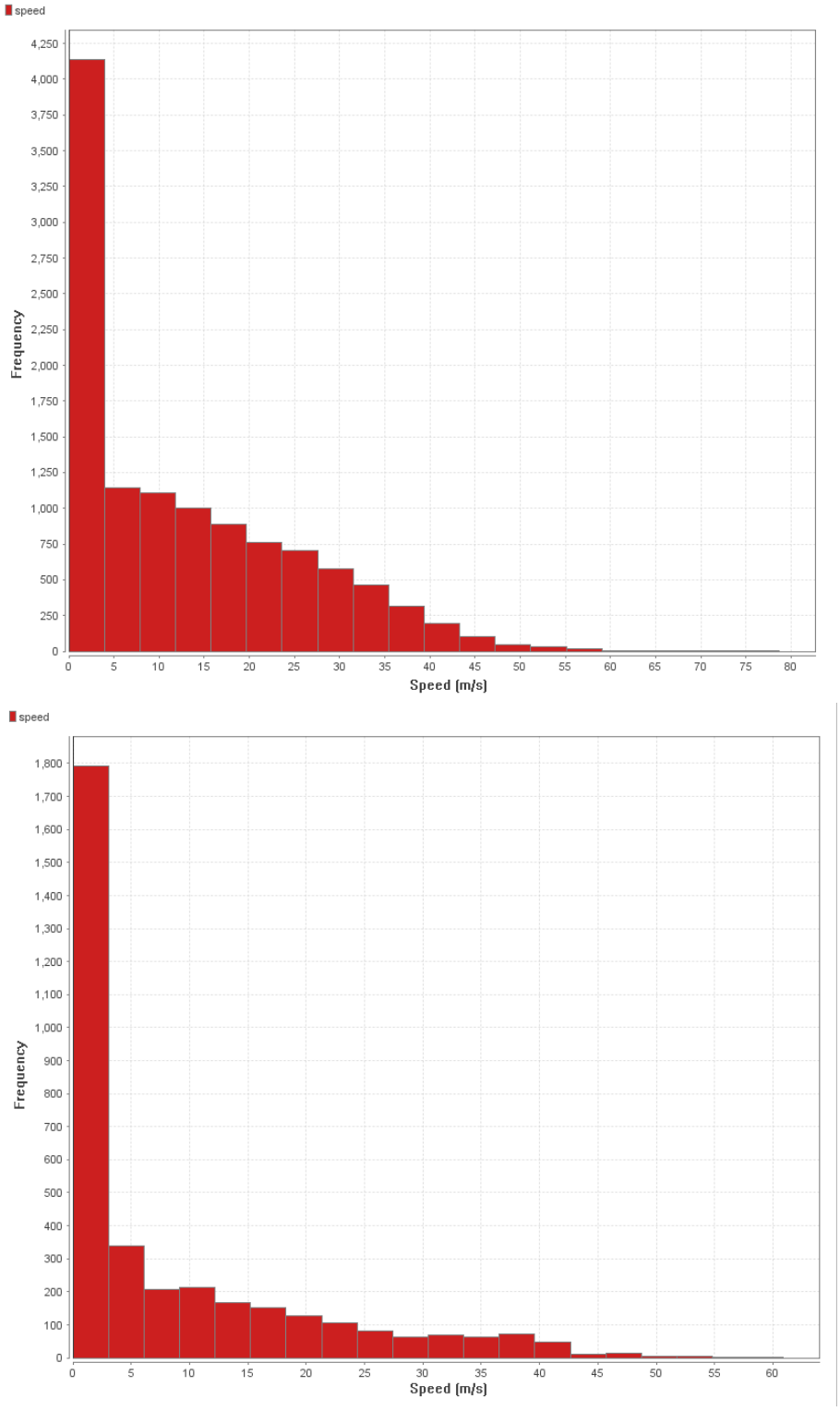


Figure 4.22: Driver speed histograms when cars were not visible (left) and when they were (right).

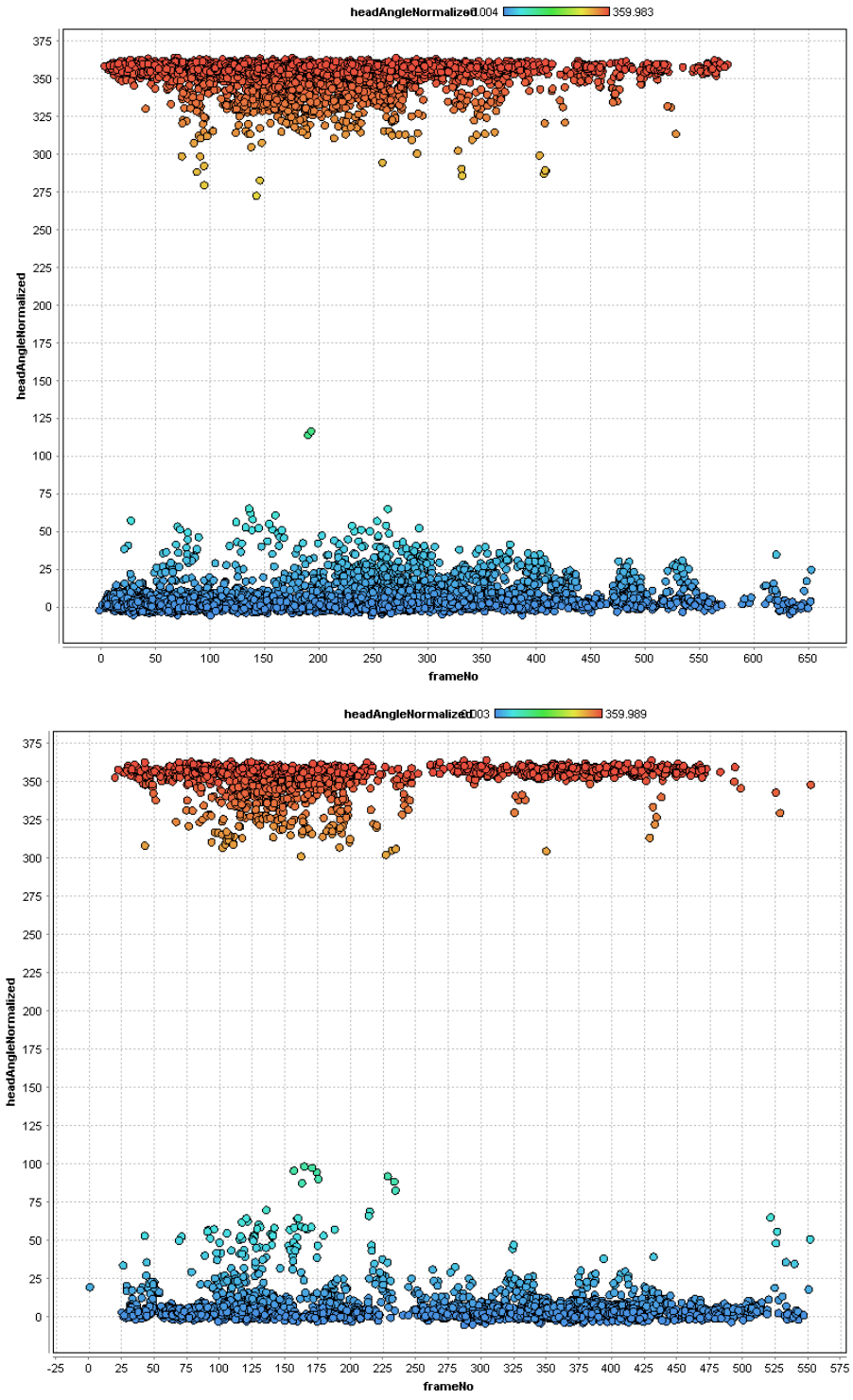


Figure 4.23: Driver head angles when cars were not visible (left) and when they were (right).

ER5: *Drivers will stop at stop signs even if there are no people incoming.*

Throughout the track several stop signs were placed in precise locations in order to test how subjects would react to them. There were intersections in which pedestrians would not cross and that featured a stop sign.

Frames in which there was a crosswalk and a stop sign visible at the same time were gathered. This subset was separated from frames where there was a crosswalk but no stop sign. A boolean value regarding the existence of the stop sign was related to the drivers' speeds at those intersections.

Results were distributed in the scatter plot in figure 4.24. Contrary to expected, drivers maintained much higher speeds in intersections where there were stop signs, and much lower speeds at ones that did not.

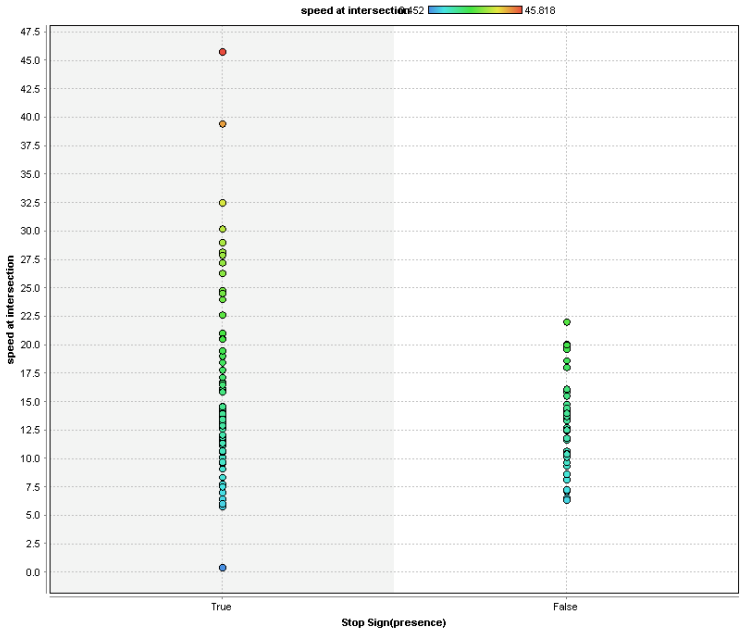


Figure 4.24: Average speed at intersections with and without a stop sign, for all intersections encountered during the experiments.

ER6: *Drivers will yield to people that made eye contact or portrayed other significant body language to them while waiting to cross the road.*

Some pedestrian encounters featured pedestrians that deliberately looked at the driver's car before crossing. They were placed in order to test drivers' actions after noticing that they want to cross through their gaze.

A process was constructed in RapidMiner Studio that separated frames where pedestrians were near a crosswalk and gazing at the car, and the driver was looking forward. Drivers' speeds and time to brake would provide insight into their thought process during this time. It was expected that drivers would be much more willing to yield after seeing a pedestrian gazing at them.

Results were compiled in the two tables in figure 4.25. Contrary to what was expected, drivers took much longer to brake after seeing the pedestrian gaze at them. However, subjects also chose to drive at significantly lower speeds in these cases (on average half the value), instead of braking.

			Min	Max	Average
speed	Real	0	0.000	48.060	15.941
firstBrakingTime	Real	0	0	33.099	3.216
speed	Real	0	0.000	38.239	7.884
firstBrakingTime	Real	0	0	29.073	10.794

Figure 4.25: Average speed and braking times for pedestrians depending on their gazing to the car. The table above is for pedestrians that didn't gaze at the car, while the bottom one is for pedestrians that did.

ER7: *Drivers will not stop as often if pedestrians that are walking towards the crosswalk take a long time.*

Some pedestrians deliberately took longer routes to the crosswalk in order to assess how the driver reacted to pedestrians that visibly wanted to cross but took a long time to do so.

From the pedestrian maps it was calculated how each of them took to change their state of *nearCrosswalk* to *crossing*. No pedestrian took longer than five seconds to reach the crosswalk, and lower times were much more frequent among pedestrians. This duration was related to the average speed and the time to brake for these scenarios.

It was expected that drivers would not stop often for pedestrians that took a long time to become an obstacle to driving. However, results in figures 4.26 and 4.27 show that drivers' actions were quite unpredictable. No single driver profile can be extracted. A big portion of drivers took little time to brake if the pedestrian also took little to time to reach he crosswalk. For longer pedestrian times, the brake time reaction varied greatly. These variations are big enough that it is not possible to make a generalization of driver behavior in these cases.

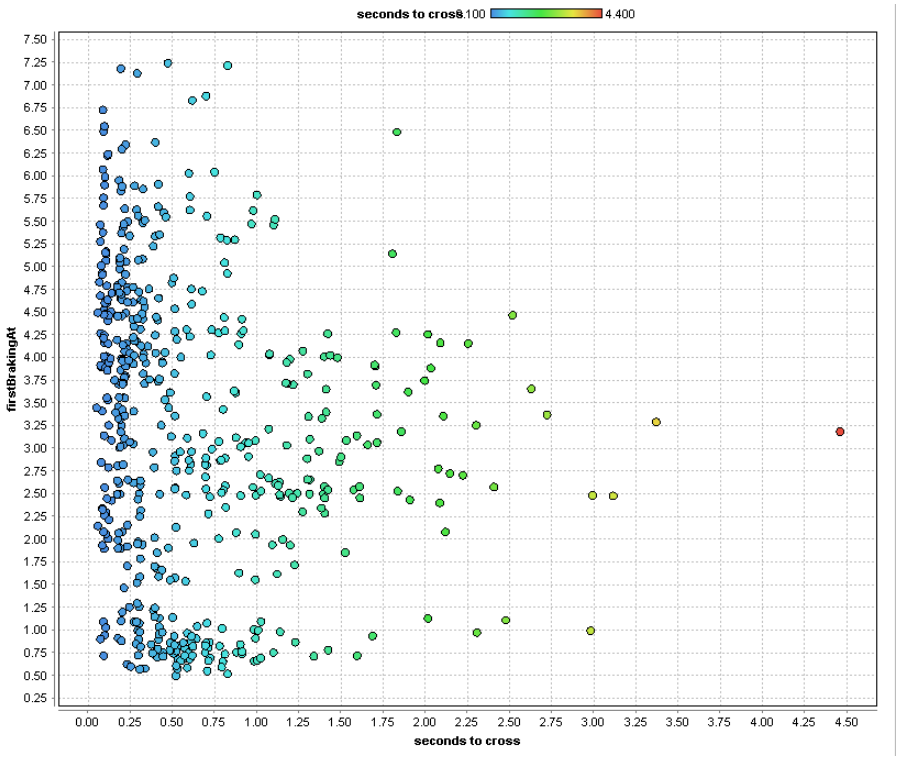


Figure 4.26: Brake times in relation to how long a pedestrian remained in a *nearCrosswalk* state.

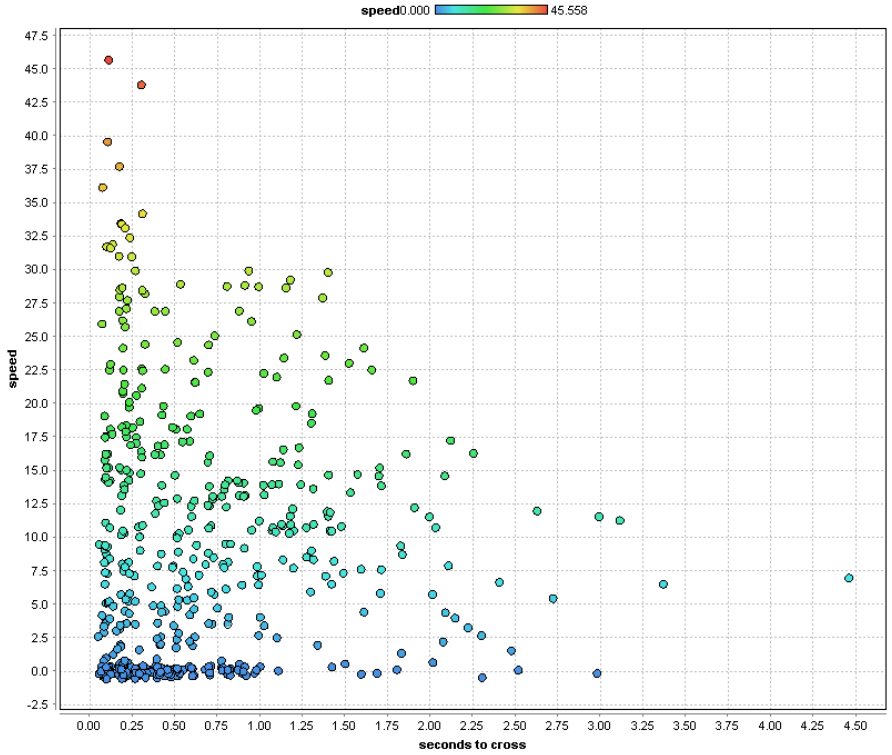


Figure 4.27: Average speed in relation to how long a pedestrian remained in a *nearCrosswalk* state.

4.2.2 Discussion

Results and their visualization provide insight into the main factors that influence drivers' actions during driver-pedestrian interactions. From the knowledge of the effect of such factors it is easier to predict the outcome of those interactions. Obtained results were compared to what was (intuitively) expected.

Starting with **ER1**, it was expected that drivers would be more willing to stop for pedestrians closer to the car. This was expected due to the influence of obstacle distance to the car. A pedestrian that is on the other end of the road evidently does not provoke the same reaction of a pedestrian only a few meters from the car. Results showed the influence of distance over drivers and how it affected their braking in encounters. A cluster model was created and visualized using RapidMiner. Three clusters were shown in order to separate the main driver profiles. From the results, it can be seen that three separate profiles show up. They are all quite distinct in what they represent to driver-pedestrian encounters. Firstly, there were the majority of drivers who decided to brake only when quite close to the pedestrian. This was the case of drivers that chose to reach the crosswalk with a relatively high speed and brake only if the pedestrian made clear that they were willing to cross. Similarly, there was another profile of drivers' behaviors. While the first cluster depicted drivers that braked quite close to the crosswalk, the opposite happened quite frequently too. These drivers maintained a large safety distance to the crosswalk, often braking at more than twice the distance than the other profile. These drivers also often maintained lower average speed throughout the whole experiment. The final cluster indicated the drivers that chose not to brake upon reaching the crosswalk. This was seen often during experiments. Some subjects simply swerved away from the incoming pedestrians, while others were clumsy with the controls and did not brake due to this. These three cluster centroids provide input into three distinct behaviors for drivers in crosswalk scenarios in relation to distance to the crosswalk and crossing pedestrians. It is important to mention that while these values show three distinct cluster profiles, the visualization and data show that for drivers that braked their behavior was not as separable as this. Drivers' distance at which they braked was rather a spectrum instead of three discrete values. Nevertheless, they might imply the presence of distinct driver profiles and their frequency to be used in a predictive tool.

In regards to **ER2**, it was expected that drivers would be more frequently yielding to pedestrians if they had not seen them for long before seeing them crossing. This was expected since the longer the pedestrian had been in sight for the driver, the better the driver could read its intentions. Pedestrians that quickly appeared and immediately crossed the road could cause misunderstandings in drivers since the interaction is so quick. Drivers' first reactions to this quick encounter should be to brake. For this analysis, the time to brake after seeing a pedestrian crossing for the first time was related to the time in which the pedestrian had already been visible to the driver. Thus, it was expected that a correlation would appear from this comparison: the longer the pedestrian had been seen the more time to brake. In fact, a correlation did seem to appear. The index shows that those factors share a correlation of 0.545. This correlation index is not very strong.

However, it is not insignificant. It shows that in some way those factors are related. Many drivers did seem to follow the pattern that was expected. However, some drivers took much less time to brake even in pedestrians seen for longer. This agrees to **ER1**'s results, for drivers that decided to brake very prematurely. Thus, this correlation could be used to have only one of these attributes be a part of the predictive tool for interactions. The other attribute could be estimated from the first one.

Regarding group size, **ER3** expected that drivers would stop more easily for pedestrians if they were accompanied by other pedestrians when crossing. Drivers should feel bigger pressure to stop if a large group of pedestrians is waiting on the side of the crosswalk. The bigger the group, the faster the driver could understand that he needed to stop. Using the group size data from the pedestrian maps a scatter plot was constructed. It related the time needed to brake after seeing the group for the first time with the size of the group. It showed that drivers chose to brake much sooner and at more consistent times if the group was bigger. It also showed that brake times were very irregular the smaller the group. For pedestrians waiting alone at the crosswalk, drivers took more liberty in their actions. For bigger groups, they braked almost always at after a short interval of time. This shows a potential negative correlation between group size and time to brake. However, it is important to say that this correlation is not perfect. For those smaller groups it is very hard to predict the brake time, whereas this is much easier for bigger groups. However, a pattern can still be found in the data. This agrees with what was expected.

The second half of experiments consisted of the same track used in the previous ones, but with cars present in the scene as well. It was expected (as per **ER4**) that these cars would influence drivers and make them driver much safer and slower. Thus, average speed was measured in frames where there were other moving cars besides the driver. It showed that in general drivers chose to drive much slower in such sections. Although slower speeds had the same general frequency in the histogram, drivers did not reach medium and high speeds as often as they did in sections with no cars. Besides this, they also moved their head much more frequently. One can infer that these choices are due to the fact that drivers had to pay attention to other cars' decisions, not only their own. Thus, other drivers are a big influence in drivers' actions. When near crosswalks their actions were quite similar, however. Since drivers were already driving much slower they did not change this behavior in crosswalks scenarios.

ER5 stated that drivers would almost always stop at stop signs, even if there were no incoming pedestrians. An analysis of frames where there were stop signs and crosswalks visible was done. It showed that in such frames drivers' actions and speeds were much more irregular than in intersections where they were non-existent. Drivers' speeds at stop signs were mostly the same if there were no signs. However, this irregularity of behaviors disproves what was expected. Contrary to what was assumed, drivers chose not to follow the stop sign rules, and continued driving regardless. Some intersections that featured stop signs did not have pedestrian triggers and thus there were no pedestrians around those areas. Moreover, some cars did not go through these areas. This rendered these intersections empty of dynamic elements. It is this lack of obstacles moving around that can be used to explain the results. Drivers felt no need to stop in those areas since there were

no threats in sight. Moreover, these sections were at the end of long straight sections in the track, which the driver usually would speed up to reach. Thus, it is wrong to assume that stop signs make no difference to drivers. It is more correct to assume that the results are the consequence of the environment that was used.

As for **ER6**, it stated that drivers would yield to people if they had made eye contact with the driver prior to crossing the road. In the real world, this is a very important protocol that pedestrians and drivers go through prior to an interaction at the crosswalk. Pedestrians may also signal explicitly that they want to cross. Drivers can't assume a pedestrian's intention as easily if their body language doesn't clarify their intentions. Therefore, it was important that this was a part of the study at hand. The results showed that in sections where pedestrians deliberately looked at drivers, the latter would drive much slower. However, they would not brake as soon. This may be because the driver noticed and acknowledged the pedestrian's presence and intention to cross, and in doing so didn't need to drive as defensively. Drivers assume that the pedestrian is willing to cross, but still attempts to bargain their yielding to pedestrians. Thus, they drive much slower but only come to a full stop much later, if pedestrian communication is maintained as assertive. This result depends on the environment that was used for testing this scenario, however. In the scenario, pedestrians looked at the car but were stood on the crosswalk. This means they conveyed different intentions through their body language. If they are stood on the crosswalk it is natural that the drivers attempt not to yield, like in all other scenarios. Nevertheless, this result is plausible, even with the constraints put upon it because of the environment used. A prediction tool would have to take the pedestrian's gaze into consideration. This gaze would also have to take into consideration the time between when the pedestrian was first visible and when it glanced at the car.

Finally, the last thing that was analyzed in the results was the influence of pedestrian time to reach the crosswalk (**ER7**). Some pedestrians in the experiments were very slow to reach the crosswalks, or showed body language that portrayed that they were indecisive to cross. Each pedestrian's time to reach a crosswalk was related to the time needed to brake and the average speed in those encounters. As the results showed, there was no correlation between any of these attributes. Most drivers maintained lowers speeds, but no generalization could be made from the results. This might mean that visibility or group size has a much bigger impact on driver judgment than the pedestrian's speed.

From all these results, it can be said that the main factors that influenced drivers were the visibility of pedestrians at the crosswalks, whether they were in a big group or by themselves, their body language (gaze) and the presence of other dynamic elements on the road. Thus, a predictive tool would make use of these factors in order to predict pedestrians' intentions. Pedestrians that are quick to appear on the drivers field of view portray a much stronger will to cross the road, more so if they are joined by others. Cars make drivers much more aware of their surroundings, and make them usually drive at slower speeds. Pedestrians that gaze at the driver immediately signal their will to cross, and this is not affected by their time to reach the crosswalk and vice-versa.

Chapter 5

Conclusions and Future Work

This chapter concludes this thesis, summarizing everything that was mentioned and alluding to possible future work to continue the study.

5.1 Conclusions

Studying driver-pedestrian interactions is not a trivial task. Many factors go into a simple encounter between a car and a pedestrian wanting to cross the road. Studies into this field are numerous, and delve into different parts of the scope of the work at hand. Collecting data about this topic can be done in many different ways, depending on the type of research study. Quantitative research may focus on visual data or instrumented car sensor data. Qualitative research may focus on questionnaires, for example. It is up to the researcher to decide the best methodology to be employed in the study. Besides this, modelling these interactions can also be done in many different ways. Most work focuses on statistical research, or the creation of simulations for quicker and more direct visualization of driver-pedestrian interactions. Finally, predicting pedestrian's intentions has to factor in many different internal and external factors of the encounter and depends on the type of data that was gathered. Recent work focused on the use of deep networks to focus on one specific attribute.

This work made use of this knowledge of the state of art to study this field. Simulations were used in order to visualize interactions more directly and repeatedly. Data was collected by instrumenting a car with a camera and sensors. This data provides visual and quantitative input that was used for analysis. Visual data had to be processed, and it was decided that image segmentation provided a simple and effective way to learn about pedestrian presence and the presence of other elements in sight. Among all data that was collected, it emphasized the study of pedestrian visibility, pedestrian eye contact, time to reach the crosswalk, other cars in sight, stop signs, pedestrian groups, among others.

Experiments were performed using twenty different subjects in a virtual environment using VR, a racing wheel and pedals. Experiments were prepared to function as a simple track where

the subject was directed from start to end, facing pedestrians in different settings along the way. Data was collected and prepared for all experiments, and was compiled in RapidMiner Studio.

Results showed that the most important factors in driver-pedestrian interactions included the pedestrians continuous visibility, their group size, other cars, among others. Results focusing on pedestrian gaze were insightful in the consequences of body language in encounters. Drivers changed their driving patters after recognizing the pedestrians' acknowledgment of their presence.

In sum, a good insight into the factors that play a role in these interactions was obtained. The methodology provides a simple and configurable way to obtain data from real subjects, and to be the basis for a predictive tool to be inserted in a simulation.

5.2 Future Work

Although some results using this methodology were satisfactory, some work to further deepen this study is necessary. It includes:

Improve the pedestrian MAS: The pedestrian agent system that was used was very hindering of the overall performance of the system. Initially, experiments were run using an environment that contained more than fifty pedestrians. However, these many pedestrians dropped performance to levels that were incompatible with the use of VR. Moreover, the pedestrian models used were limited, and a deeper study that took into consideration the pedestrians' demographics could have been done.

Improve the car MAS: The car MAS used was an external tool that integrated well with Unity. However, during simulations, the car agents were extremely slow. Although their speed was configurable, it could not rise over a certain amount, and it was still very low. This meant that any deeper research into car influence on drivers could not be done. A better, more dynamic car system would mean a deeper study of other drivers' actions over the subject.

Use other types of segmentation techniques: The segmentation was performed using an external Unity library. If data was to be gathered using real-life car visual data, a different segmentation method would be required. Mask-RCNNs could suffice in this role, but later their integration into Unity is not trivial.

Study pedestrian gestures and pose: Pedestrian body language is much more than their eye contact with the driver. Their posture and gestures can convey their intentions perhaps more clearly than eye contact. Thus, a deeper study using poses would be necessary. Mask-RCNNs could also fulfill this role, although performance and integration would have to be factored into their use.

Intention prediction: All data and methodology in order to setup a predictive tool were put in place. RapidMiner Studio could allow for quick setup of such a tool. It would be crucial to create a new data set for validation and testing of predictions. This tool could be implanted into the car agents in order to simulate real driving using knowledge learned from human drivers, thus contributing to a more realistic and dynamic environment.

References

[AGR+13] P. R. J. A. Alves, J. Gonçalves, R. J. F. Rossetti, E. C. Oliveira, and C. Olaverri-Monreal. Forward collision warning systems using heads-up displays: Testing usability of two new metaphors. In *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–6, June 2013.

[arc07] ARCHISIM: a behavioural multi-actors traffic simulation model for the study of a traffic system including ITS aspects. Viol. 5(1):7–16, 2007.

[ARF+14] J.E. E Almeida, R.J.F. J F Rossetti, B.M. M Faria, J.T. T Jacob, and A.L. L Coelho. Towards a methodology for human behaviour elicitation: Preliminary results. *26th European Modeling and Simulation Symposium, EMSS 2014*, pages 220–228, 2014.

[ARJ+17] João Emílio Almeida, Rosaldo J F Rossetti, João Tiago Pinheiro Neto J.T.P.N. Jacob, Brígida Mónica Faria, A Le?a Coelho, Rosaldo J F Rossetti, and António Leça Coelho. Serious games for the human behaviour analysis in emergency evacuation scenarios. *Cluster Computing*, 20(1):707–720, 2017.

[BK09] Cornelia Boenisch and Tobias Kretz. Simulation of pedestrians crossing a street. 11 2009.

[BZSMM13] Marilyne Brosseau, Sohail Zangenehpour, Nicolas Saunier, and Luis Miranda-Moreno. The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in montreal. *Transportation Research Part F: Traffic Psychology and Behaviour*, 21:159 – 172, 2013.

[CCB+18] Fanta Camara, Serhan Cosar, Nicola Bellotto, Natasha Merat, and Charles W. Fox. Towards pedestrian-av interaction: method for elucidating pedestrian preferences. 10 2018.

[CGM+18] Fanta Camara, Oscar Giles, Ruth Madigan, M Rothmüller, P Holm Rasmussen, SA Vendelbo-Larsen, Gustav Markkula, Yee Mun Lee, Laura Garach, Natasha Merat, and CW Fox. Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. July 2018.

[Chu12] Yishih Chung. Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident; analysis and prevention*, 61, 09 2012.

[CPD+18] R Currano, So Yeon Park, Lawrence Domingo, Jesus Garcia-Mancilla, Pedro Santana, Victor Gonzalez, and Wendy Ju. ¡vamos!: Observations of pedestrian interactions with driverless cars in mexico. pages 210–220, 09 2018.

[DHGRH80] Kenneth D. Hopkins, Gene Glass, and B R. Hopkins. Basic statistics for the behavioral sciences. *Technometrics*, 22, 05 1980.

[DRC+17] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. 11 2017.

[DSD+17] Shuchisnigdha Deb, Lesley Strawderman, Janice Dubien, Brian Smith, Daniel Carruth, and Teena Garrison. Evaluating pedestrian behavior at crosswalks: Validation of a pedestrian behavior questionnaire for the u.s. population. *Accident Analysis Prevention*, 106:191–201, September 2017.

[DT17] N. Deo and M. M. Trivedi. Learning and predicting on-road pedestrian behavior around vehicles. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6, October 2017.

[Dí02] Emilio Moyano Díaz. Theory of planned behavior and pedestrians' intentions to violate traffic regulations. *Transportation Research Part F*, 5:169–175, April 2002.

[EREHJW86] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back propagating errors. *Nature*, 323:533–536, 10 1986.

[FC+18] C W Fox, F Camara, et al. When should the chicken cross the road ?: Game theory for autonomous vehicle - human interactions. *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, pages 431–439, 2018.

[FL18] Zhijie Fang and Antonio M. López. Is the pedestrian going to cross? answering by 2d pose estimation. *CoRR*, abs/1807.10580, 2018.

[FRBR09] M. C. Figueiredo, R. J. F. Rossetti, R. A. M. Braga, and L. P. Reis. An approach to simulate autonomous vehicles in urban traffic scenarios. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6, Oct 2009.

[GCVB18] Andrea Gorrini, Luca Crociani, Giuseppe Vizzari, and Stefania Bandini. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59:269–285, 11 2018.

[GDDM13] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 11 2013.

[GRJ+14] J. S. V. Gonçalves, R. J. F. Rossetti, J. Jacob, J. Gonçalves, C. Olaverri-Monreal, A. Coelho, and R. Rodrigues. Testing advanced driver assistance systems with a serious-game-based human factors analysis suite. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 13–18, June 2014.

[GRO12] J. Gonçalves, R. J. F. Rossetti, and C. Olaverri-Monreal. Ic-deep: A serious games based application to assess the ergonomics of in-vehicle information systems. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 1809–1814, Sep. 2012.

[HFJS06] Julie Hatfield, Ralston Fernandes, R. F. Soames Job, and Ken Smith. Misunderstanding of right-of-way rules at various pedestrian crossing types : Observational study and survey. *Accident Analysis and Prevention*, 39:833–842, December 2006.

[HL+18] Azra Habibovic, Victor Malmsten Lundgren, et al. Communicating intent of automated vehicles to pedestrians. *Frontiers in Psychology*, 9, August 2018.

[Hoo02] Serge P. Hoogendoorn. Extracting microscopic pedestrian characteristics from video data results from experimental research into pedestrian walking behavior. volume 1, pages 1–15, 2002.

[HS06] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science (New York, N.Y.)*, 313:504–7, 08 2006.

[Kit17] Genshiro Kitagawa. Information criteria for statistical modeling in data-rich era. *ECONVN 2018: Econometrics for Financial Applications*, pages 20–43, December 2017.

[KV13] Raghuram Kadali and P Vedagiri. Modelling pedestrian road crossing behaviour under mixed traffic condition. *European Transport - Trasporti Europei*, 55, December 2013.

[LD15] Nils Lubbe and Johan Davidsson. Drivers' comfort boundaries in pedestrian crossings: A study in driver braking characteristics as a function of pedestrian walking speed. *Safety Science*, 75:100–106, 06 2015.

[LLZX08] Xiugang Li, Dominique Lord, Yunlong Zhang, and Yuanchang Xie. Predicting motor vehicle crashes using support vector machine models. *Accident Analysis Prevention*, 40(4):1611 – 1618, 2008.

[LT06] Tsippy Lotan and Tomer Toledo. In-vehicle data recorder for evaluation of driving behavior and safety. *Transportation Research Record*, 1953:112–119, 01 2006.

[LTWT14] P. Luo, Y. Tian, X. Wang, and X. Tang. Switchable deep network for pedestrian detection. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 899–906, June 2014.

[MNA16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. pages 565–571, 10 2016.

[MS16] Emese Mako and Petra Szakonyi. Evaluation of human behaviour at pedestrian crossings. *Transportation Research Procedia*, 14:2121–2128s, 2016.

[MSSE15] S. V. Mamidipalli, V. Sisiopiku, B. Schroeder, and L. Elefteriadou. A review of analysis techniques and data collection methods for modeling pedestrian crossing behaviors. *Journal of Multidisciplinary Engineering Science and Technology*, 2(2), February 2015.

[NS04] Wassim G. Najm and David L. Smith. Modeling driver response to lead vehicle decelerating. In *SAE Technical Paper*. SAE International, March 2004.

[PR12] José L. F. Pereira and Rosaldo J. F. Rossetti. An integrated architecture for autonomous vehicles simulation. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, SAC '12, pages 286–292, New York, NY, USA, 2012. ACM.

[PXP00] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000. PMID: 11701515.

[RAKG13] Rosaldo J.F. Rossetti, Joao Emilio Almeida, Zafeiris Kokkinogenis, and Joel Goncalves. Playing transportation seriously: Applications of serious games to artificial transportation systems. 28:107,113, July 2013.

[RCMS18] Vasili Ramanishka, Yi-Ting Chen, Teruhisa Misu, and Kate Saenko. Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning. *CoRR*, abs/1811.02307, 2018.

[Ris85] Ralf Risser. Behavior in traffic conflict situations. *Accident; analysis and prevention*, 17:179–97, May 1985.

[RKT17] Amir Rasouli, Yulia Kotseruba, and John Tsotsos. Agreeing to cross: How drivers and pedestrians communicate. pages 264–269, 06 2017.

[RL15] Rosaldo J.F. Rossetti and Ronghui Liu, editors. *Advances in Artificial Transportation Systems and Simulation*. Academic Press, Boston, 2015.

[RLT11] R. J. F. Rossetti, R. Liu, and S. Tang. Guest editorial special issue on artificial transportation systems and simulation. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):309–312, June 2011.

[RM14] Lior Rokach and Oded Maimon. *Data Mining With Decision Trees: Theory and Applications*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2nd edition, 2014.

[SAPY17] Ahmad Sallab, Mohammed Abdou, Etienne Perot, and Senthil Yogamani. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, 2017:70–76, January 2017.

[SAW17] Khaled Shaaban and Karim Abdel-Warith. Agent-based modeling of pedestrian behavior at an unmarked midblock crossing. *Procedia Computer Science*, 109:26–33, 2017.

[SMM] Khaled Shaaban, Deepti Muley, and Abdulla Mohammed. Analysis of illegal pedestrian crossing behavior on a major divided arterial road.

[SNL03] David L. Smith, Wassim G. Najm, and Andy H. Lam. Analysis of braking and steering performance in car-following scenarios. In *SAE Technical Paper*. SAE International, 03 2003.

[SR11] J. Schroeder Schroeder and Nagui M. Rouphail. Event-based modeling of driver yielding behavior at unsignalized crosswalks. *Journal of Transportation Engineering*, 137:455–465, July 2011.

[SSL05] Armin Seyfried, Bernhard Steffen, and Thomas Lippert. Basics of modelling the pedestrian flow. *Physica A: Statistical Mechanics and its Applications*, 368, July 2005.

[SUBTW02] Dazhi Sun, Satish Ukkusuri, Rahim Benekohal, and S. Travis Waller. Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. *Urbana*, 51, 12 2002.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.

[SZW+15] Rouxian Sun, Xiangling Zhuang, Changxu Wu, Guozhen Zhao, and Kan Zhang. The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 30:97–106, April 2015.

[Var98] Andras Varhelyi. Drivers' speed behaviour at a zebra crossing: A case study. *Accident; analysis and prevention*, 30:731–43, 12 1998.

[Wer06] Paul Werbos. *Backwards Differentiation in AD and Neural Nets: Past Links and New Opportunities*, volume 50, pages 15–34. 02 2006.

[Wol16] Ingo Wolf. *The Interaction Between Humans and Autonomous Agents*, pages 103–124. 05 2016.

[WSB15] Gennady Waizman, Shraga Shoval, and Itzhak Benenson. Micro-simulation model for assessing the risk of vehicle-pedestrian road accidents. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 19:63–77, 2015.

[WW10] D. Wanty and S. M. Wilkie. Trialling pedestrian countdown timers at traffic signals. *Wellington: NZ Transport Agency Research Report 428*, 2010.

[YLW16] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Pedestrian behavior understanding and prediction with deep neural networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 263–279, 2016.

[YS14] Ambar Yadav and Arti Singh. Driving simulator. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 16:33–38, May 2014.

[ZLT16] Zijian Zheng, Pan Lu, and Denver Tolliver. Decision tree approach to accident prediction for highway–rail grade crossings: Empirical analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2545:115–122, 01 2016.

[Šu14] Matús Šucha. Road users' strategies and communication: driver-pedestrian interaction. *Transport Research Arena (TRA) 2014 Proceedings*, 2014.

REFERENCES

Appendix A

Consent form

The following page includes the consent form (based on the Helsinki declaration) given out and explained to subjects in the experiments. As explained in the experiment protocol, subjects were explained the details of the experiments, as well as the setup and the goal of the experiment. No subjects chose to abandon the experience. In total, twenty subjects were gathered and signed the consent form.

The consent form states that my authorship of this thesis and the conduction of the experiments, as well as a declaration that the subject understood the conditions of the experiment. Besides that, it acknowledges that the subject knows the possible discomfort during the experiment, as well as the lack of consequences of it. Finally, it states that the subject is able to leave the experiment at any time, and requests his consent in the study, formalizing it with a signature.

DECLARAÇÃO DE CONSENTIMENTO

(Baseada na declaração de Helsínquia)

No âmbito da realização da tese de Mestrado do Mestrado Integrado em Engenharia Informática da Faculdade de Engenharia da Universidade do Porto, intitulada **Simulating Driver-Pedestrian Interaction and Intentions Inference**, realizada pelo estudante **Rui Pedro Correia Soares**, orientada pelo Prof. João Jacob e sob a co-orientação do Prof. Rosaldo Rossetti, eu abaixo assinado, _____, declaro que compreendi a explicação que me foi fornecida acerca do estudo em que irei participar, nomeadamente o carácter voluntário dessa participação, tendo-me sido dada a oportunidade de fazer as perguntas que julguei necessárias.

Tomei conhecimento de que a informação ou explicação que me foi prestada versou os objectivos, os métodos, o eventual desconforto e a ausência de riscos para a minha saúde, e que será assegurada a máxima confidencialidade dos dados.

Explicaram-me, ainda, que poderei abandonar o estudo em qualquer momento, sem que daí advenham quaisquer desvantagens.

Por isso, consinto participar no estudo e na recolha de imagens necessárias, respondendo a todas as questões propostas.

Porto, 3 de Junho de 2019

_____

(Participante ou seu representante)

Appendix B

Comparison of average speeds in frames with and without pedestrians

An analysis of sections in the experiments where there were pedestrians and its comparison to sections where there were no pedestrians was done in section 4. Figure 4.17 shows the comparison of the gathered data in a histogram format. It was mentioned that the average speed was lower for sections where there were pedestrians. This was done via a statistical comparison of the means of the two populations, from the gathered samples.

Let population **1** be the speed values in frames where there are no pedestrians. Similarly, let population **2** be the speed values in frames where there are pedestrians.

Let n_p be the size of the sample of population p.

$$n_1 = n_2 = 37$$

Let $\bar{x_p}$ be the mean of the sample gathered of population p.

$$\bar{x_1} = 15.177164$$
$$\bar{x_2} = 9.406347$$

These values represent the estimated means of each of the populations. This gives us no concrete information over the actual difference of the means of the populations. Let σ_p be the standard deviation of population p and let s_p be the standard deviation of the sample of population p.

$$\bar{s_1} = 4.760785$$
$$\bar{s_2} = 1.961947$$

The difference of the means of the samples can give us a starting point in calculating the

difference of the means of the populations.

$$\bar{x_1} - \bar{x_2} = 5.770817$$

A confidence interval of 95% is desired. Thus:

$$\alpha = 1 - 0.95 = 0.05$$

The z-value for this alpha corresponds to:

$$z_{\alpha/2} = z_{0.025} = 1.96$$

Finally, the difference of the means of the two populations is defined as such:

$$(\bar{x_1} - \bar{x_2}) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Replacing the variables with their values:

$$5.770817 \pm 1.96 \sqrt{\frac{4.760785^2}{37} + \frac{1.961947^2}{37}} =$$

$$= 5.770817 \pm 1.659187$$

Thus, with 95% confidence we can say that the difference of the means of the average speeds in frames with and without pedestrians is between:

$$[4.11, 7.43]$$

This value is quite significant in the urban scenarios in context.